IDS #12

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, C12N 5/10, A61K 48/00, G01N 33/68, C12Q 1/68, C07K 16/18

A3

(11) International Publication Number: WO 95/34649

(43) International Publication Date: 21 December 1995 (21.12.95)

(21) International Application Number: PCT/GB95/01386

(22) International Filing Date: 13 June 1995 (13.06.95)

(30) Priority Data:

9411900.5	14 June 1994 (14.06.94)	GB
PCT/GB94/02822	23 December 1994 (23.12.94)	GB
9507766.5	13 April 1995 (13.04.95)	GB
08/422,582	14 April 1995 (14.04.95)	US

(71) Applicants *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB). LEIDEN UNIVERSITY [NL/NL]; P.O. Box 9500, NL-2300 RA Leiden (NL). UNIVERSITY OF WALES COLLEGE OF MEDICINE [GB/GB]; Heath Park, Cardiff CF4 4XN (GB). ERASMUS UNIVERSITY ROTTERDAM [NL/NL]; Burg Ondlaan 50, P.O. Box 1738, NL-3000 DR Rotterdam (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HARRIS, Peter, Charles [GB/GB]; 65 Freelands Road, Oxford OX4 4BS (GB). PERAL, Belen [ES/GB]; 77 Lock Crescent, Kidlington, Oxford OX5 1HF (GB). WARD, Christopher, James [GB/GB]; 30 Benson Road, Oxford OX3 7EH (GB). HUGHES, James [GB/GB]; 225 Cowley Road, Oxford OX4 1XD (GB). BREUNING, Martin, Hendrik [NL/NL]; Brigantijnstraat 57, NL-1503 BR Zaandam (NL). PETERS, Dorothea, Johanna, Maria [NL/NL]; Zuster Meijboomstraat 267, NL-2331 PH Leiden (NL). ROELFSEMA, Jeroen, Hendrik [NL/NL]; Vijf Meilaan 2006, NL-2321 RR Leiden (NL). SAMPSON, Julian [GB/GB]; 34 Bridge Street, Cardiff CF5 2EL (GB). HALLEY, Dirkje, Jorijntje, Johanna [NL/NL]; Van Aerssenlaan 35 d, NL-3039 KD Rotterdam (NL). NELLIST, Mark, David [GB/NL]; Noordmolenstraat 57b, NL-3053 RG Rotterdam (NL). JANSSEN, Lambertus, Antonius, Jacobus [NL/NL]; Schokker 37, NL-2991 DJ Barendrecht (NL). HESSELING, Arjenne, Ligue, Wilhelma [NL/NL]; Haya van Someren Downerpad 7, NL-3207 DK Spijkenisse (NL).

(74) Agents: NEWELL, William, Joseph et al.; Wynne-Jones, Laine & James, 22 Rodney Road, Cheltenham, Gloucestershire GL50 1JJ (GB).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) Date of publication of the international search report:
4 January 1996 (04.01.96)

(54) Title: POLYCYSTIC KIDNEY DISEASE 1 GENE AND USES THEREOF

(57) Abstract

The present invention relates to the polycystic kidney disease 1 (PKD1) gene and its nucleic acid sequence, mutations thereof in patients having PKD1-associated disorders, the protein encoded by the PKD1 gene or its mutants, and their uses in disease diagnosis and therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/GB 95/01386

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/47 C12N5/10 A61K48/00 G01N33/68
C12Q1/68 C07K16/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N A61K C12Q C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	J. AM. SOC. NEPHROL., vol. 4, no. 3, page 814 G. GERMINO ET AL 'A novel approach to the identification of the PKD1 gene'	1-4,6-30
Y	see abstract 91p	31-40

Y	KIDNEY INTERNATIONAL, vol. 43, no. supp 3, 19 May 1993 pages s20-s25, G. GERMINO ET AL 'Positional cloning approach to the dominant polycystic kidney disease gene, PKD1' see the whole document	1-40
	---/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
28 November 1995	05.12.1995

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Van der Schaal, C

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GENOMICS, vol. 13, pages 144-151, G. GERMINO ET AL 'The gene for autosomal dominant polycystic kidney disease.....' cited in the application see the whole document especially page 150, left column, last paragraph ---	1-40
Y	A. GRIFFITHS ET AL 'An introduction to genetic analysis' , W. FREEMAN AND COMPANY , NEW YORK see page 427 see page 453, left column, last paragraph - right column, paragraph 1 see page 453, right column, last paragraph - page 461 ---	1-40
A	CURRENT OPINION IN GENETICS AND DEVELOPMENT, vol. 3, pages 425-431, J. MULLEY ET AL 'Integrating maps of chromosome 16' ---	
X	EMBL DATABASE, Accession no. T04943 sequence reference HS9431, August 30, 1993 M. ADAMS et al, '3400 Expressed sequence tags identify diversity of transcripts from human brain & NATURE GENETICS, vol. 4, 1993 pages 256-267, ---	1-4,12
X	EMBL DATABASE, Accession no T05931 sequence reference HS9312, August 30, 1993, M. ADAMS et al., '3400 Expressed sequnce tags identify diversity of transcripts from human brain & NATURE GENETICS, vol. 4, 1993 pages 256-267, ---	1-4,12
X	EMBL DATABASE, Accession no T09245 sequnce reference T09245, August 7 1993 M. ADAMS et al 'Rapid cDNA sequencing (expresserd sequence tags).... & NATURE GENETICS, vol. 4, 1993 pages 373-380, --- -/--	1-4,12

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/GB 95/01386

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	CELL, vol. 77, 17 June 1994 pages 881-894, C. WARD ET AL 'The polycystic kidney disease 1 gene encodes a 14kb transcript and lies within a duplicated region on chromosome 16' see the whole document	1,3-5,7, 9-15, 17-31, 33-40
E,L	WO,A,95 18225 (MEDICAL RES COUNCIL ;LEIDEN UNIVERSITY (NL); UNIV WALES MEDICINE () 6 July 1995 see the whole document	1,3-5,7, 9-15, 17-31, 33-40
P,X	HUM. MOL. GENET. (1995), 4(4), 575-82 CODEN: HMGEE5;ISSN: 0964-6906, April 1995 BURN, TIMOTHY C. ET AL 'Analysis of the genomic sequence for the autosomal dominant polycystic kidney disease (PKD1) gene predicts the presence of a leucine-rich repeat' see the whole document	1,2, 5-16, 20-40
P,X	CELL (CAMBRIDGE, MASS.) (1995), 81(2), 289-98 CODEN: CELLB5;ISSN: 0092-8674, 21 April 1995 GLUECKSMANN-KUIS, M. ET AL 'Polycystic kidney disease: the complete structure of the PKD1 gene and its protein' see the whole document	1,2, 5-16, 20-40
T	CELL (CAMBRIDGE, MASS.) (1995), 81(7), 1171 CODEN: CELLB5;ISSN: 0092-8674, 1995 'The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. [Erratum to document cited in CA121:150194]' see paragraph 1	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

national application No.
PCT/GB95/01386

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 31-33
because they relate to subject matter not required to be searched by this Authority, namely:
Remark: Although claims 31-33 are directed to a method of treatment of the human body, the search has been carried out and based on the alleged effect of the compound.

2. [ ] Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searches without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int ional Application No
PCT/GB 95/01386

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9518225	06-07-95	AU-B-	1322695	17-07-95
		AU-B-	1322795	17-07-95
		WO-A-	9518226	06-07-95

PATENT & TRADEMARK OFFICE DEPOSIT ACCOUNT CHARGE(S)

Patent No.: 09/548,882 **Date of Charge:** December 18, 2002

Client Matter No. 23893-7083
(AQ 2017.00)

	Narrative for Fees	Amount
	Filing Fee for filing Provisional Application	$
	Surcharge-Late Filing of Declaration	$
	Assignment Recordation Fee	$
	Petition for Extension of Time: **(Small Entity)** ☐ 1 ☐ 2 ☐3 ☐ 4 ☐ 5 Months	$
	Additional Claims independent..... total claims multiple dependent.....	$ $ $
	Notice of Appeal (Small Entity)	$
	Issue Fee	$
	PCT Filing and Related Fees	
	Chapter II Demand Fee	
	Other: Charge for Petition to Revive Application	$ 640.00
	TOTAL	$640.00

*Please reimburse the Patent Office Deposit Account No.: **50-2518** for the total amount indicated.*

Approved By: _____ January 6, 2002
(Sign Name) Date

Antoinette Konski _____ 7491 _____
(Print Name) Attorney Number

Nichols, Peggy

From: CMS EVENT SERVER
Sent: Wednesday, January 01, 2003 10:01 AM
To: Nichols, Peggy
Subject: Daily Notification of Cash Receipts

Yesterday we received payment for the matters and invoices listed below. If there is a balance remaining on the invoice, the amount is shown. If there are any questions about this notification, please forward this e-mail with your questions to your Cash Receipts Coordinator - Willie Belin or Diana Medina.

****

THIS MESSAGE WAS GENERATED AUTOMATICALLY. DO NOT REPLY DIRECTLY TO THIS MESSAGE. RESPONSES ARE NOT MONITORED. THANK YOU.

Billing Attorney:	#7491, Konski, Antoinette F.
C/M#:	24460.7001
Client:	Sidney Kimmel Cancer Center
Matter:	US-Methods for Classifying Tumors Accord
Invoice #:	508957
Invoice Date:	9/30/2002
Invoice Amount:	$26,130.33
Payment Received From:	Bingham McCutchen trustee
Payment Date:	12/31/2002
Payment Amount:	$15,000.00
Invoice Balance, all Payors:	$11,130.33

****

THIS MESSAGE WAS GENERATED AUTOMATICALLY. DO NOT REPLY DIRECTLY TO THIS MESSAGE. RESPONSES ARE NOT MONITORED. THANK YOU.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, C07K 14/47, C12N 5/10, A61K 48/00, G01N 33/68, C12Q 1/68, C07K 16/18

A2

(11) International Publication Number: WO 95/34649

(43) International Publication Date: 21 December 1995 (21.12.95)

(21) International Application Number: PCT/GB95/01386

(22) International Filing Date: 13 June 1995 (13.06.95)

(30) Priority Data:

9411900.5	14 June 1994 (14.06.94)	GB
PCT/GB94/02822	23 December 1994 (23.12.94)	GB
9507766.5	13 April 1995 (13.04.95)	GB
08/422,582	14 April 1995 (14.04.95)	US

(71) Applicants *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB). LEIDEN UNIVERSITY [NL/NL]; P.O. Box 9500, NL-2300 RA Leiden (NL). UNIVERSITY OF WALES COLLEGE OF MEDICINE [GB/GB]; Heath Park, Cardiff CF4 4XN (GB). ERASMUS UNIVERSITY ROTTERDAM [NL/NL]; Burg Ondlaan 50, P.O. Box 1738, NL-3000 DR Rotterdam (NL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: HARRIS, Peter, Charles [GB/GB]; 65 Freelands Road, Oxford OX4 4BS (GB). PERAL, Belen [ES/GB]; 77 Lock Crescent, Kidlington, Oxford OX5 1HF (GB). WARD, Christopher, James [GB/GB]; 30 Benson Road, Oxford OX3 7EH (GB). HUGHES, James [GB/GB]; 225 Cowley Road, Oxford OX4 1XD (GB). BREUNING, Martin, Hendrik [NL/NL]; Brigantijnstraat 57, NL-1503 BR Zaandam (NL). PETERS, Dorothea, Johanna, Maria [NL/NL]; Zuster Meijboomstraat 267, NL-2331 PH Leiden (NL). ROELFSEMA, Jeroen, Hendrik [NL/NL]; Vijf Meilaan 2006, NL-2321 RR Leiden (NL). SAMPSON, Julian [GB/GB]; 34 Bridge Street, Cardiff CF5 2EL (GB). HALLEY, Dirkje, Jorijntje, Johanna [NL/NL]; Van Aerssenlaan 35 d, NL-3039 KD Rotterdam (NL). NELLIST, Mark, David [GB/NL]; Noordmolenstraat 57b, NL-3053 RG Rotterdam (NL). JANSSEN, Lambertus, Antonius, Jacobus [NL/NL]; Schokker 37, NL-2991 DJ Barendrecht (NL). HESSELING, Arjenne, Ligue, Wilhelma [NL/NL]; Haya van Someren Downerpad 7, NL-3207 DK Spijkenisse (NL).

(74) Agents: NEWELL, William, Joseph et al.; Wynne-Jones, Laine & James, 22 Rodney Road, Cheltenham, Gloucestershire GL50 1JJ (GB).

(81) Designated States: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: POLYCYSTIC KIDNEY DISEASE 1 GENE AND USES THEREOF

(57) Abstract

The present invention relates to the polycystic kidney disease 1 (PKD1) gene and its nucleic acid sequence, mutations thereof in patients having PKD1-associated disorders, the protein encoded by the PKD1 gene or its mutants, and their uses in disease diagnosis and therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

POLYCYSTIC KIDNEY DISEASE 1 GENE AND USES THEREOF

BACKGROUND TO THE INVENTION

In humans, one of the commonest of all genetic disorders is autosomal dominant polycystic kidney disease (ADPKD) also termed adult polycystic kidney disease (APKD), affecting approximately 1/1000 individuals (Dalgaard, 1957). ADPKD is a progressive disease of cyst formation and enlargement typically leading to end stage renal disease (ESRD) in late middle age. The major cause of morbidity in ADPKD is progressive renal disease characterized by the formation and enlargement of fluid filled cysts, resulting in grossly enlarged kidneys. Renal function deteriorates as normal tissue is compromised by cystic growth, resulting in end stage renal disease (ESRD) in more than 50% of patients by the age of 60 years (Gabow, et al., 1992). ADPKD accounts for 8-10% of all renal transplantation and dialysis patients in Europe and the USA (Gabow, 1993).

ADPKD also causes cystic growth in other organs (reviewed in Gabow, 1990) and occasionally presents in childhood (Fink, et al., 1993; Zerres, et al., 1993). Extrarenal manifestations include liver cysts (Milutinovic, et al., 1980), and more rarely cysts of the pancreas (Gabow, 1993) and other organs. Intracranial aneurysms occur in approximately 5% of patients and are a significant cause of morbidity and mortality due to subarachnoid haemorrhage (Chapman, et al., 1992). ADPKD is associated with a higher prevalence of various connective tissue disorders. An increased prevalence of heart valve defects (Hossack, et

al., 1988), hernia (Gabow, 1990) and colonic diverticulae (Scheff, et al., 1980) have been reported.

Considerable progress has been made in the last few years in understanding the pathophysiology of ADPKD (and other animal models of cystic disease). Cysts in ADPKD are known to develop from outpouchings of descending or ascending kidney tubules and the early stages are characterized by a thickening and disorganization of the basement membrane, accompanied by a de-differentiation of tubular epithelial cells. Several of the characteristics of ADPKD epithelia: altered growth responses, abnormal expression of various proteins and reversal of polarity, may be a sign of this de-differentiation and important in cyst expansion. The nature of the primary defect which triggers these changes is, however, unknown and consequently much effort has been devoted to identifying the causative agent by genetic means.

The first step towards positional cloning of an ADPKD gene was the demonstration of linkage of one locus now designated the polycystic kidney disease 1 (PKD1) locus to the α globin cluster on the short arm of chromosome 16 (Reeders, et al., 1985). Subsequently, families with ADPKD unlinked to markers of 16p were described (Kimberling, et al., 1988; Romeo, et al., 1988) and a second ADPKD locus (PKD2) has recently been assigned to chromosome region 4q13-q23 (Kimberling, et al., 1993; Peter, et al., 1993). It is estimated that approximately 85% of ADPKD is due to PKD1 (Peters and Sankuijl, 1992) with PKD2 accounting for most of

the remainder. PKD2 appears to be milder condition with a later age of onset and ESRD (Parfrey. et al., 1990; Gabow, et al., 1992; Ravine, et al., 1992).

The position of the PKD1 locus was refined to chromosome band 16p13.3 and many markers were isolated from that region (Breuning, et al., 1987; Reeders, et al., 1988; Breuning, et al., 1990; Germino, et al., 1990; Hyland, et al., 1990; Himmelbauer, et al., 1991). Their order, and the position of the PKD1 locus, has been determined by extensive linkage analysis in normal and PKD1 families and by the use of a panel of somatic cell hybrids (Reeders et al., 1988; Breuning, et al., 1990; Germino, et al., 1990). ADPKD is genetically heterogenous with loci mapped not only to 16p13.3 (PKD1), but also to chromosome 4 (PKD2). Although the phenotype of PKD1 and PKD2 are clearly similar, it is now well documented that PKD1 (which accounts for about 85% of ADPKD; (Peters, 1992) is a more severe disease with an average age at ESRD of about 56 years compared to about 71.5 years for PKD2 (Ravine, 1992). An accurate long range restriction map of the 16p13.3 region (Harris, et al., 1990; Germino, et al., 1992) has located the PKD1 locus in an interval of approximately 600 kb between the markers GGG1 and SM7 (Harris, et al., 1991; Somlo, et al., 1992) (see Figure 1a). The density of CpG islands and identification of many mRNA transcripts indicated that this area is rich in gene sequences. Germino et al. (1992) estimated that the candidate region contains approximately 20 genes.

Identification of the PKD1 gene from within this area

has thus proved difficult and other means to pinpoint the disease gene have been sought. Linkage disequilibrium has been demonstrated between PKD1 and the proximal marker VK5, in a Scottish population (Pound, et al., 1992) and between PKD1 and BLu24 (see Figure 1a), in a Spanish population (Peral, et al., 1994). Studies with additional markers have shown evidence of a common ancestor in a proportion of each population (Peral, et al., 1994; Snarey, et al., 1994), but the association has not precisely positioned the PKD1 locus.

Disease associated genomic rearrangements, detected by cytogenetics or pulsed field gel electrophoresis (PFGE) have been instrumental in the identification of various genes associated with various genetic disorders. Hitherto, no such abnormalities related to PKD1 have been described. This situation contrasts with that for the tuberous sclerosis locus, which lies within 16p13.3 (TSC2). In that case, TSC associated deletions were detected by PFGE within the interval thought to contain the PKD1 gene and their characterisation was a significant step toward the rapid identification of the TSC2 gene (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). The TSC2 gene therefore maps within the candidate region for the hitherto unidentified PKD1 gene; as polycystic kidneys are a feature common to TSC and ADPKD1 (Bernstein and Robbins, 1991) the possibility of an etiological link, as proposed by Kandt et al. (1992), was considered. A contiguous gene syndrome resulting from the disruption of PKD1 and the adjacent tuberous sclerosis 2 (TSC2) gene, which is associated with

TSC and severe childhood onset polycystic kidney disease, has also been defined (Brook-Carter et al, 1994).

We have now identified a pedigree in which the two distinct phenotypes, typical ADPKD or TSC, are seen in different members. In this family, the two individuals with ADPKD are carriers of a balanced chromosome translocation with a breakpoint within 16p13.3. We have located the chromosome 16 translocation breakpoint and a gene disrupted by this rearrangement has been defined; the discovery of additional mutations of that gene in other PKD1 patients shows that we have identified the PKD1 gene. Full characterisation of the PKD1 transcript has been significantly complicated because of the unusual genomic region containing most of the gene. All but 3.5 kb at the 3' end of the transcript (which is about 14 kb in total) is encoded by a region which is reiterated several times elsewhere on the same chromosome (in 16p13.1 and termed the HG area). The structure of the duplication is complex, with some regions copied more times than others, and the HG region encoding three large transcripts. The transcripts from the HG area are: HG-A (21 kb), HG-B (17 kb) and HG-C (8.5 kb) and although these have 3' ends which differ from PKD1, over most of their length they share substantial homology to the PKD1 transcript. Consequently, cloning and characterizing a bona fide PKD1 cDNA has proven difficult. To overcome the problem caused by duplication we have cloned cDNAs covering the entire transcript from a cell line which contains the PKD1 but not the HG loci. Characterisation of

these cDNAs has enabled the PKD1 protein sequence to be predicted and led to the identification of several homologies with described motifs.

SUMMARY OF THE INVENTION

Accordingly, in one aspect, this invention provides an isolated, purified or recombinant nucleic acid sequence comprising:

(a) a PKD1-encoding nucleic acid or its complementary strand,

(b) a sequence substantially homologous to, or capable of hybridizing to, a substantial portion of a molecule defined in (a) above, or

(c) a fragment of a molecule defined in (a) or (b) above.

In particular, there is provided a sequence wherein the PKD1 gene has the nucleic acid sequence according to Fig. 15, or the partial sequence of Figs. 7 or 10. The invention therefore includes a DNA molecule coding for a polypeptide having the amino acid sequence of Figure 15, or a polypeptide fragment thereof; and genomic DNA corresponding to a molecule as in (a) - (c) above.

As used herein, "substantially homologous" refers to a nucleic acid strand that is sufficiently duplicative of the PKD1 sequence presented in Fig. 15 such that it is capable of hybridizing to that sequence under moderately stringent, and preferably stringent conditions, as defined herein below. Preferably, "substantially homologous" refers to a homology of between 97 and 100%. Further, such a strand will encode or be complementary to a strand that encodes PKD1 protein having the biological activity described below. As used herein, a "substantial portion of a molecule" refers

to at least 60%, preferably 80% and most preferably 90% of the molecule in terms of its linear residue length or its molecular weight. "Nucleic acid" refers to both DNA and RNA.

The PKD1 gene described herein is a gene found on human chromosome 16, and the results of studies described herein form the basis for concluding that this PKD1 gene encodes a protein called PKD1 protein which has a role in the prevention or suppression of ADPKD. The PKD1 gene therefore includes the DNA sequences shown in Figure 15, and all functional equivalents. By "functional equivalents", we mean nucleic acid sequences that are substantially homologous to the PKD1 nucleic acid sequence, as presented in Fig. 15, and encoding a protein that possesses one or more of the biological functions or activities of PKD1; i.e., that is involved in cell/cell adhesion, cell/cell recognition or cell/cell communication, for example to effect adhesion of cells to other cells or components of the extracellular matrix; effect communication and/or interaction between epithelial cells and the basal membrane (whether in kidneys or otherwise); assist in development of connective tissue such as assembly and/or maintenance of the basal membrane; in signal transduction between cells or cells and components of the extracellular matrix; and/or to promote binding of cells carrying proteins such as integrins or carbohydrates to target cells. The biological function of PKD1 of course includes maintaining a healthy physiological state; that is, the native protein's

aberrations or absence results in ADPKD or an associated disorder.

The PKD1 gene may furthermore include regulatory regions which control the expression of the PKD1 coding sequence, including promoter, enhancer and terminator regions. Other DNA sequences such as introns spliced from the end-product PKD1 RNA transcript are also encompassed. Although work has been carried out in relation to the human gene, the corresponding genetic and functional sequences present in lower animals are also encompassed.

The present invention therefore further provides a PKD1 gene or its complementary strand having the sequence according to Figure 15 which gene or strand is mutated in some ADPKD patients (more specifically, PKD1 patients). Therefore, the invention further provides a nucleic acid sequence comprising a mutant PKD1 gene as described herein, including wherein Intron 43 as defined hereinbelow has a deletion of 18 or 20bp resulting in an intron of 55 or 57bp.

As used herein, "PKD1 mutant" or "mutation" encompasses alterations of the native PKD1 nucleotide or amino acid sequence, as defined by Fig. 15, i.e., substitutions, deletions or additions, and also encompasses deletion of DNA containing the entire PKD1 gene.

The invention further provides a nucleic acid sequence comprising a mutant PKD1 gene, especially one selected from a sequence comprising a partial sequence according to Figures 7 and/or 10, or the corresponding sequences disclosed in Fig. 15, when:

(a) [OX114] base pairs 1746-2192 as defined in Figure 7 are deleted (446bp);

(b) [OX32] base pairs 3696-3831 as defined in Figure 7 are deleted by a splicing defect;

(c) [OX875] about 5.5kb flanked by the two XbaI sites shown in Figure 3a are deleted and the EcoR1 site separating the CW10 (41kb) and JH1 (18kb) sites is thereby absent

(d) [WS53] about 100kb extending between the JH1 and CW21 and the SM6 and JH17 sites shown in Figure 6 and the PKD1 gene is thereby absent, the deletion lying proximally between SM6 and JH13;

(e) [461] 18bp are deleted in the 75bp intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(f) [OX1054] 20bp are deleted in the 75bp intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(g) [WS212] about 75kb are deleted between SM9-CW9 distally and the PKD1 3'UTR proximally as shown in Figure 12;

(h) [WS-215] about 160kb are deleted between CW20 and SM6-JH17 as shown in Figure 12;

(i) [WS-227] about 50kb are deleted between CW20 and JH11 as shown in Figure 12;

(j) [WS-219] about 27kb are deleted between JH1 and JH6 as shown in Figure 12;

(k) [WS-250] about 160kb are deleted between CW20 and Blu24 as shown in Figure 12;

(1) [WS-194] about 65kb is deleted between CW20 and CW10.

The invention therefore extends to RNA molecules comprising an RNA sequence corresponding to any of the DNA sequences set out above. Such molecule may be the transcript reference PBP and identifiable with respect to the restriction map of Figure 3a and having a length of about 14 KB.

In another aspect, the invention provides a nucleic acid probe having a sequence as set out above; in particular, this invention extends to a purified nucleic acid probe which hybridizes to at least a portion of the DNA or RNA molecule of any of the preceding sequences. Preferably, the probe includes a label such as a radiolable, for example, a ^{32}P label.

In another aspect, this invention provides a purified DNA or RNA coding for a protein comprising the amino acid sequence of Figure 15, or a protein polypeptide having homologous properties with said protein, or having at least one functional domain or active site in common with said protein.

The DNA molecule defined above may be incorporated in a recombinant cloning vector for expressing a protein having the amino acid sequence of Figure 15, or a protein or a polypeptide having at least one functional domain or active site in common with said protein. Such a vector may include any vector for expression in bacteria, e.g., E. coli; yeast, insect, or mammalian cells.

The invention also features a nucleic acid probe for detecting PKD1 nucleic acid comprising 10 consecutive nucleotides as presented in Fig. 15. Preferably, the probe may comprise 15, 20, 50, 100, 200, or 300, etc., consecutive nucleotides (nt) presented in Fig. 13, and may fall within the size range 15nt-13kb, 100nt-5kb, 150nt-4kb, 300nt-2kb, and 500nt-1kb.

Probes are used according to the invention in hybridization reactions to identify PKD1 sequences, whether they be native or mutated PKD1 DNA or RNA, as disclosed herein. Such probes are useful for identifying the PKD1 gene or a mutation thereof, as defined herein.

The invention also features a synthetic polypeptide corresponding in amino acid residue sequence to at least a portion of the sequence of naturally occurring PKD1, and having a molecular weight equal to less than that of the native protein. A synthetic polypeptide of the invention is useful for inducing the production of antibodies specific for the synthetic polypeptide and that bind to naturally occurring PKD1.

Preferred embodiments of this aspect of the invention include a group of synthetic polypeptides whose members correspond to a fragment of the PKD1 protein comprising a stretch of amino acids of at least 8, and preferably 15, 30, 50, or 100 residues in length from the sequence disclosed in Fig. 15.

In another aspect, the invention provides a polypeptide encoded by a sequence as set out above, or having the amino

acid sequence according to the amino acid sequence of Figure 15, or a protein or polypeptide having homologous properties with said protein, or having at least one functional domain or active site in common with said protein. In particular, there is provided an isolated, purified or recombinant polypeptide comprising a PKD1 protein or a mutant or variant thereof or encoded by a sequence set out above or a variant thereof having substantially the same activity as the PKD1 protein. The present invention may further comprise a polypeptide having 9 or 13 transmembrane pairs instead of 11 transmembrane domains as described hereinbelow. Further comprising this invention is a molecule which interacts with a polypeptide as herein described which molecule synergises, causes, enhances or is necessary for the functioning of the PKD1 protein as herein described.

The invention also encompasses recombinant expression vectors comprising a nucleic acid or isolated DNA encoding PKD1 and a process for preparing PKD1 polypeptide, comprising culturing a suitable host cell comprising the vector under conditions suitable for promoting expression of PKD1, and recovering said PKD1.

This invention also provides an <u>in vitro</u> method of determining whether an individual is at risk of a PKD1-associated disorder, comprising assaying a biological sample from the individual to determine the presence and/or amount of PKD1 protein or polypeptide having the amino acid sequence of Figure 15.

As used herein, "biological sample" includes any fluid

or tissue sample from a mammal, preferably a human, including but not limited to blood, urine, saliva, any body organ tissue, cells from any body tissue, including blood cells.

Additionally or alternatively, a sample may be assayed to determine the presence and/or amount of mRNA coding for the protein or polypeptide having the amino acid sequence of Figure 15, or to determine the fragment lengths of fragments of nucleotide sequences coding for the protein or polypeptide of Figure 15, or to detect inactivating mutations in DNA coding for a protein having the amino acid sequence of Figure 15 or a protein having homologous properties. The screening preferably includes applying a nucleic acid amplification process, as described herein in detail, to said sample to amplify a fragment of the DNA sequence. The nucleic acid amplification process advantageously utilizes at least one of the following sets of primers as identified herein: AH3 F9 : AH3 B7; 3A3 C1 : 3A3 C2; and AH4 F2 : JH14 B3.

Alternatively, the screening method may comprise digesting the sample DNA to provide EcoRI fragments and hybridizing with a DNA probe which hybridizes to the EcoRI fragment identified (A) in Figure 3(a), and the DNA probe may comprise the DNA probe CW10 identified herein.

Another screening method may comprise digesting the sample to provide BamHI fragments and hybridizing with a DNA probe which hybridizes to the BamHI fragment identified (B) in Figure 3(a), and the DNA probe may comprise the DNA probe

1A1H.6 identified herein.

A method according to the present invention may comprise detecting a PKD1-associated disorder in a patient suspected of having or having predisposition to the disorder (i.e., a carrier), the method comprising detecting the presence of and/or evaluating the characteristics of PKD1 DNA, PKD1 mRNA and.or PKD1 protein in a sample taken from the patient. Such method may comprise detecting and/or evaluating whether the PKD1 DNA is deleted, missing, mutated, aberrant or not expressing normal PKD1 protein. One way of carrying out such a method comprises: A. taking a biological, tissue or biopsy sample from the patient; B. detecting the presence of and/or evaluating the characteristics of PKD1 DNA, PKD1 mRNA and/or PKD1 protein in the sample to obtain a first set of results; C.comparing the first set of results with a second set of results obtained using the same or similar methodology for an individual that is not suspected of having the disorder; and if the first and second sets of results differ in that the PKD1 DNA is deleted, missing, aberrant, mutated or not expressing PKD1 protein then that is indicative of the presence, predisposition or tendency of the patient to develop the disorder. As used herein, a "PKD1-associated disorder" refers to adult polycystic kidney disease, as described herein, and also refers to tuberous sclerosis, as well as other disorders having symptoms such as cyst formation in common with these diseases.

A specific method according to the invention comprises

extracting from a patient a sample of PKD1 DNA or DNA from the PKD1 locus purporting to be PKD1 DNA, cultivating the sample in vitro and analyzing the resulting protein, and comparing the resulting protein with normal PKD1 protein according to the well-established Protein Truncation Test. Less sensitive tests include analysis of RNA using RT PCR (reverse transcriptase polymerase chain reaction), and examination of genomic DNA.

Step C of the above method may be replaced by: comparing the first set of results with a second set of results obtained using the same or similar methodology in an individual that is known to have the or at least one of the disorder(s); and if the first and second sets of results are substantially identical, this indicates that the PKD1 DNA in the patient is deleted, mutated or not expressing normal PKD1 protein.

The invention further provides a method of characterizing a mutation in a subject suspected of having a mutation in the PKD1 gene, which method comprises:A amplifying each of the exons in the PKD1 gene of the subject; B. denaturing the complementary strands of the amplified exons; C.diluting the denatured separate, complementary strands to allow each single-stranded DNA molecule to assume a secondary structural confirmation; D. subjecting the DNA molecule to electrophoresis under non-denaturing conditions; E. comparing the electrophoresis pattern of the single-stranded molecule with the electrophoresis pattern of a single-stranded molecule

containing the same amplified exon from a control individual which has either a normal or PKD1 heterozygous genotype; and, F. sequencing any amplification product which has an electrophoretic pattern different from the pattern obtained from the DNA of the control individual.

The invention also extends to a diagnostic kit for carrying out a method as set out above, comprising nucleic acid primers for amplifying a fragment of the DNA or RNA sequences defined above, and packaging means therefore. The kit may optionally include written instructions stating that the primers are to be used for detection of disorders associated with the PKD1 gene. The nucleic acid primers may comprise at least one of the following sets: AH3 F9 : AH3 B7; 3A3 C1 : 3A3 C2; and AH4 E2 : JH14 B3.

Another embodiment of kit may combine one or more substances for digesting a sample to provide EcoRI fragments and a DNA probe as previously defined. A further embodiment of kit may comprise one or more substances for digesting a sample to provide BamHI fragments and a DNA probe as previously defined.

A vector (such as Bluescript (available from Stratagene)) comprising a nucleic acid sequence set out above; and a host cell (such as E. coli strain SL-1 Blue (available from Stratagene)) transfected or transformed with the vector are also provided, together with the use of such a vector or a nucleic acid sequence set out above in gene therapy and/or in the preparation of an agent for treating or preventing a PKD1-associated disorder.

Therefore, there is further provided a method of treating or preventing a PKD1-associated disorder which method comprises administering to a patient in need thereof a functional PKD1 gene to affected cells in a manner that permits expression of PKD1 protein therein and/or a transcript produced from a mutated chromosome (such as the deleted WS-212 chromosome) which is capable of expressing functional-PKD1 protein therein.

As used herein, the term "hybridization" refers to conventional DNA/DNA or DNA/RNA hybridization conditions. For example, for a DNA or RNA probe of about 10 - 50 nucleotides, moderately stringent hybridization conditions are preferred and include 10X SSC, 5X Denhardts, 0.1% SDS, at 35 - 50 degrees for 15 hours; for a probe of about 50 - 300 nucleotides, "stringent" hybridization conditions are preferred and refer to hybridization in 6X SSC, 5X Denhardts, 0.1% SDS at 65 degrees for 15 hours.

The present invention further provides the use of PKD1 protein or polycystin or a mutant or variant thereof having substantially the same biological activity there as in therapy. In particular, to effect cell adhesion, recognition or communication for example to effect adhesion of cells to other cells or components of the extracellular matrix; effect communication and/or interaction between epithelial cells and the basal membrane (whether in kidneys or otherwise); assisting in development of connective tissue such as assembly and/or maintenance of the basal membrane; in signal transduction between cells or cells and components

of the extracellular matrix; and/or to promote binding of cells carrying proteins such as integrins or carbohydrates to target cells.

Accordingly, where it is preferred to administer the polypeptide directly to a patient in need thereof, the invention further provides the use of a PKD1 protein or polycystin in the preparation of a medicament. Therefore, there is also provided a pharmaceutical formulation comprising a PKD1 protein, functional PKD1 gene and/or a transcript produced from a mutated chromosome which is capable of expressing functional PKD1 protein, in association with a pharmaceutically acceptable carrier therefor.

The invention also features an immunoglobin, i.e., a polyclonal or monoclonal antibody specific for an epitope of PKD1, which epitope is found in the amino acid sequence presented in Fig. 15.

The invention also features a method of assaying for the presence of PKD1 in a sample of mammalian, preferably human cells, comprising the steps of: (a) providing an antibody specific for said PKD1; and (b) assaying for the presence of PKD1 by admixing an aliquot from a sample of mammalian cells with antibody under conditions sufficient to allow for formation and detection of an immune complex of PKD1 and the antibody. Such method is useful for detecting disorders involving aberrant expression of the PKD1 gene or processing of the protein, as described herein.

Preferably, this method includes providing a monoclonal

antibody specific for an epitope that is antigenically the same, as determined by Western blot assay, ELISA or immunocytochemical staining, and substantially corresponds in amino acid sequence to the amino acid sequence of a portion of PKD1 and having a molecular weight equal to less than that of PKD1.

The invention thus also features a kit for detecting PKD1, the kit including at least one package containing an antibody or idiotype-containing polyamide portion of an antibody raised to a synthetic polypeptide of this invention or to a conjugate of that polypeptide bound to a carrier. An indicating group or label is utilized to indicate the formation of an immune reaction between the antibody and PKD1 when the antibody is admixed with tissue or cells.

Further features will become more fully apparent in the following description of the embodiments of this invention and from the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Before describing preferred embodiments of the invention in detail, the drawings will briefly be described.

Figure 1a (top): A long range map of the terminal region of the short arm of chromosome 16 showing the PKD1 candidate region defined by genetic linkage analysis. The positions of selected DNA probes and microsatellites used for haplotype, linkage or heterozygosity analyses are indicated. Markers previously described in linkage disequilibrium studies are shown in bold (from: Harris, et al., 1990; Harris, et al., 1991; Germino, et al., 1992; Somlo, et al., 1992; Peral, et al., 1994; Snarey, et al., 1994).

(bottom): A detailed map of the distal part of the PKD1 candidate region showing: the area of 16p13.3 duplicated in 16p13.1 (hatched); C, Cla I restriction sites; the breakpoints in the somatic cell hybrids, N-OH1 and P-MWH2A; DNA probes and the TSC2 gene. The limits of the position of the translocation breakpoint found in family 77 (see b), determined by evidence of heterozygosity (in 77-4) and PFGE (see c and text) is also indicated. The contig covering the 77 breakpoint region consists of the cosmids: 1, CW9D; 2, ZDS5; 3, JH2A; 4, REP59; 5, JC10.2B; 6, CW10III; 7, SM25A; 8, SMII; 9, NM17.

Figure 1b: Pedigree of family 77 which segregates a 16;22 translocation; showing the chromosomal composition of each subject. Individuals 77-2 and 77-3 have the balanced products of the exchange - and have PKD1; 77-4 is monosomic

for 16p13.3-->16pter and 22q11.21-->22pter - and has TSC.

Figure 1c: PFGE of DNA from members of the 77 family: 77-1 (1); 77-2 (2); 77-3 (3); 77-4 (4); digested with Cla I and hybridised with SM6. In addition to the normal fragments of 340 and partially digested fragment of 480 kb a proximal breakpoint fragment of approximately 100 kb (arrowed) is seen in individuals, 77-2, 77-3 and 77-4; concordant with segregation of the der(16) chromosome.

Figure 2: FISH of the cosmid CW10III (cosmid 6; Figure 1a) to a normal male metaphase. Duplication of this locus is illustrated with two sites of hybridisation on 16p; the distal site (the PKD1 region) is arrowed. The signal from the proximal site (16p13.1) is stronger than that from the distal, indicating that sequences homologous to CW10III are reiterated in 16p13.1.

Figure 3a: A detailed map of the 77 translocation region showing the precise localisation of the 77 breakpoint and the region that is duplicated in 16p13.1 (hatched). DNA probes (open boxes); the transcripts, PKD1 and TSC2 (filled boxes; with direction of transcription indicated by an arrow) and cDNAs (grey boxes) are shown below the genomic map. The known genomic extent of each gene is indicated at the bottom of the diagram and the approximate genomic locations of each cDNA is indicated under the genomic map. The positions of genomic deletions found in PKD1 patients, OX875 and OX114, are also indicated. Restriction sites for EcoR I (E) and incomplete maps for BamH I (B); Sac I (S) and Xba I (X) are shown. SM3 is a 2kb BamH1 fragment shown at

the 5' end of the gene.

Figure 3b: Southern blots of BamH I digested DNA from individuals: 77-1 (1); 77-2 (2); and 77-4 (4) hybridised with: left panel, 8S3 and right panel, 8S1 (see a). 8S3 detects a novel fragment on the telomeric side of the breakpoint (12 kb: arrowed) associated with the der(22) chromosome in 77-2, but not 77-4; 8S1 identifies a novel fragment on the centromeric side of the breakpoint (9 kb: arrowed) - associated with the der(16) chromosome - in 77-2 and 77-4. The telomeric breakpoint fragment is also seen weakly with 8S1 (arrowed) indicating that the breakpoint lies in the distal part of 8S1. The 8S3 and 8S1 loci are both duplicated; the normal BamH I fragment detected at the 16p13.3 site by these probes is 11 kb (see a), but a similar sized fragment is also detected at the 16p13.1 site. Consequently, the breakpoint fragments are much fainter than the normal (16p13.1 plus 16p13.3) band.

Figure 4a: PBP cDNA, 3A3, hybridised to a Northern blot containing about 1 µg polyA selected mRNA per lane of the tissue specific cell lines: lane 1, MJ, EBV-transformed lymphocytes; lane 2, K562, erythroleukemia; lane 3, FS1, normal fibroblasts; lane 4, HeLa, cervical carcinoma; lane 5, G401, renal Wilm's tumour; lane 6, Hep3B, hepatoma; lane 7, HT29, colonic adenocarcinoma; lane 8, SW13, adrenal carcinoma; lane 9, G-CCM, astrocytoma. A single transcript of approximately 14 kb is seen; the highest level of expression is in fibroblasts and in the astrocytoma cell line, G-CCM. Although in this comparative experiment little

expression is seen in lanes 1, 4 and 7, we have demonstrated at least a low level of expression in these cell lines on other Northern blots and by RT-PCR (see later).

Figure 4b: A Northern blot containing about 20 µg of total RNA from the cell line G-CCM hybridised with cDNAs or a genomic probe which identify various parts of the PBP gene. Left panel, a single about 14 kb transcript is seen with a cDNA from the single copy area, 3A3. Right panel, a cDNA, 21P.9, that is homologous to parts of the region that is duplicated (JH12, JH8 and JH10; see Figure 3a) hybridises to the PBP transcript and three novel transcripts; HG-A (about 21 kb), HG-B (about 17 kb) and HG-C (8.5 kb). A similar pattern of transcripts is seen with cDNAs and genomic fragments that hybridise to the area between JH5 and JH13, with the exception of the JH8 area. Middle panel, JH8 hybridises to the transcripts PBP, HG-A and HG-B but not to HG-C.

Figure 4c: A Northern blot of 20 µg total fibroblast RNA from: normal control (N); 77-2 (2); 77-4 (4) hybridised with 8S1, which contains the 16;22 translocation breakpoint (see Figure 3). A transcript of about 9 kb (PBP-77) is identified in the two patients with this translocation but not in the normal control. PBP-77 is a chimeric PBP transcript formed due to the translocation and is not seen in 77-2 or 77-4 RNA with probes which map distal to the breakpoint.

Figure 5a: FIGE of DNA from: normal (N) and ADPKD patient OX875 (875), digested with EcoR I and hybridised

with, left panel, CW10; middle panel, JH1. Normal fragments of 41 kb (plus a 31 kb fragment from the 16p13.1 site), CW10, and 18 kb, JHI, are identified with these probes; OX875 has an additional 53 kb band (arrowed). The EcoR I site separating these two fragments is removed by the deletion (see Figure 3a). The right panel shows a Southern blot of BamH I digested DNA (as above) hybridised with 1A1H.6. A novel fragment of 9.5 kb is seen in OX875 DNA, as well as the normal 15 kb fragment. These results indicate that OX875 has a 5.5 kb deletion; its position was determined more precisely by mapping relative to two Xba I sites which flank the deletion (see figure 3a).

Figure 5b: Northern blot of total fibroblast RNA, as (a), hybridised with the cDNAs, AH4, 3A3 and AH3. A novel transcript (PBP-875) of about 11 kb is seen with AH4 (the band is reduced in intensity because the probe is partly deleted) and AH3 (arrowed), which flank the deletion, but not 3A3 which is entirely deleted (see figure 3a). The transcripts HG-A, HG-B and HG-C, from the duplicated area, are seen with AH3 (see figure 4b).

Figure 5c: Left panel; FIGE of DNA from: normal (N) and ADPKD patient OX114 (114), digested with EcoR I and hybridised with CW10; a novel fragment of 39 kb (arrowed) is seen in OX114. Middle panel; DNA, as above, plus the normal mother (M) and brother (B) of OX114 digested with BamH I and hybridised with CW21. A larger than normal fragment of 19 kb (arrowed) was detected in OX114 but not other family members due to deletion of a BamH I site; together these

results are consistent with a 2 kb deletion (see Figure 3a). Right panel; RT-PCR of RNA, as above, with primers flanking the OX114 deletion (see Experimental Procedures). A novel fragment of 810 bp (arrowed) is seen in OX114, indicating a deletion of 446 bp in the PBP transcript.

Figure 5d: RT-PCR of RNA from: ADPKD patient OX32 (32) plus the probands, normal mother (M) and affected father (F) and sibs (1) and (2) using the C primer pair from 3A3 (see Experimental Procedures). A novel fragment of 125 bp is detected in each of the affected individuals.

Figure 6: Map of the region containing the TSC2 and PBP genes showing the area deleted in patient WS-53 and the position of the 77 translocation breakpoint. Localisation of the distal end of the WS-53 deletion was described (European Chromosome 16 Tuberous Sclerosis Consortium, 1993) and we have now localised the proximal end between SM6 and JH17. The size of the aberrant Mlu I fragment in WS-53, detected by JH1 and JH17, is 90kb and these probes lie on adjacent Mlu I fragments of 120kb and 70kb, respectively. Therefore the WS-53 deletion is about 100kb. Restriction sites for: Mlu I (M); Nru I (R); Not I (N); and partial maps for Sac II (S) and BssH II (H) are shown. DNA probes (open boxes) and the TSC2 and PBP transcripts (filled boxes) are indicated below the line with their known genomic extents (brackets). The locations of the microsatellites KG8 and SM6 are also indicated.

Figure 7: The partial nucleotide sequence (cDNA) of the PKD1 transcript extending 5631bp to the 3' end of the

gene. The corresponding predicted protein (also shown in SEQ ID NO: 4:) is shown below the sequence and extends from the start of the nucleotide sequence. The GT-repeat, KG8, is in the 3' untranslated region between 5430-5448 bp. This sequence corresponds to GenBank Accession No. L33243 and is shown in SEQ ID NO: 3:.

Figure 8: The sequence of the probe 1A1H0.6 (also shown in SEQ ID NO: 5:).

Figure 9: The sequence (SEQ ID NO: 6:) of the probe CW10 which is about 0.5kb.

Figure 10: The larger partial nucleotide sequence (SEQ ID NO: 1:) of the PKD1 transcript (cDNA) extending from bp 2 to 13807bp to the 3' end of the gene together with the corresponding predicted protein (also shown in SEQ ID NO: 2:). This larger partial sequence encompasses the (smaller) partial sequence of Figure 7 from amino acid no. 2726 in SEQ ID NO: 3: and relates to the entire PKD1 gene sequence apart from its extreme 5' end.

Figure 11: A map of the 75bp intron amplified by the primer set 3A3C insert at position 3696 of the 3' sequence showing the positions of genomic deletions found in PKD1 patients 461 and OX1054.

Figure 12: A map of the region of chromosome 16 containing the TSC2 and PKD1 genes showing the areas affected in patients WS-215, WS-250, WS-212, WS-194, WS-227 and WS-219; also WS-53 (but cf. Figure 6). Genomic sites for the enzymes MluI (M), ClaI (C), PvuI (P) and NruI (R) are shown. Positions of single copy probes and cosmids used

to screen for deletions are shown below the line which represents about 400kb of genomic DNA. The genomic distribution of the approximately 45kb TSC2 gene and known extent of the PKD1 gene are indicated above. The hatched area represents an about 50kb region which is duplicated more proximally on chromosome 16p.

Figure 13 is a genomic map of the PKD1 gene. (Top) A restriction map of the genomic area containing the PKD1 gene showing sites for Bam H1(B), EcoRI(E) and partial maps for XbaI (X) and Hind III(H), and the duplicated area (hatched). The position of genomic clones and the cosmid JH2A are shown above the map (open boxes). The positions of the 46 exons of the PKD1 gene are shown below the map (solid boxes, translated areas; open boxes, untranslated regions; UTRs). Each 5th exon is numbered and the direction of transcription arrowed. The area sequenced in Figs. 7 and 10 is bracketed and the approximate location of the 3' end of the TSC2 gene is shown on the left (dashed line and hatched box). (Bottom) The cDNA contig covering the PKD1 transcript. The cDNAs are: 1, rev1; 2, S13;3, S3/4; 4, S1/3;5, GAP e; 6, GAP d; 7, GAP g; 8, GAP a (see table 2 for details); 9, A1C; 10, AH3; 11, 3A3; 12, AH4.

Figure 14 (a) (Top): Map of the genomic BamH I fragment, SM3 which contains the CpG island at the 5' end of the PKD1 gene, showing the probe CW45 (open box). Genomic restriction sites for the methylation sensitive enzymes: SacII (S), NotI (N), MluI (M) and BssHII (H) are illustrated. The approximate position of the DNaseI

hypersensitive site is also shown (large arrow), plus the location of the first exon including the proposed transcription start site (small arrow), the 5'UTR (open box) and the translated region (solid bar). (Bottom) The GC content across the area is plotted with a window size of 50 nt. A peak of GC content of over 80% is seen in the area of the transcriptional start site and the first exon. A corresponding lack of CpG suppression was also found with an average CpG/GC ratio of 0.84 between 800-1,800 bp.

Figure 14(b). Analysis of DNase I hypersensitivity at the PKD1 CpG island. DNA isolated from HeLa cells treated with an increasing amount of DNase I (left to right: first lane contains no DNase 1), digested with BamH I and hybridised with CW45. A fragment about 400 bp smaller than the restriction fragment is seen with increasing DNase 1, indicating a hypersensitive site as shown in (a). SM3 is within the duplicated area and so both the PKD1 and HG loci are assayed together. The degree of DNaseI digestion seen at the end of the assay indicates that cleavage occurs at the PKD1 and HG loci.

Figure 15 provides the sequence of the PKD1 transcript and predicted protein. The full sequence of 14,148 bp from the transcription start site to the poly A tail is shown. The probable signal sequence of 23 amino acids is shown after the first methionine (underlined) plus the cleavage site (arrow). The predicted transmembrane (TM) domains (double underlined and numbered) and N-linked glycosylation sites (asterisk) are indicated. The position of a possible

hinge sequence is underlined and tyrosine kinase and protein kinase C phosphorylation sites marked with a box and circle, respectively.

Figure 16(a). The leucine rich repeats (LRRs) found in the PKD1 protein (72-125aa) are compared with each other and to the LRR consensus (Rothberg, 1990; Kobe, 1994); a, aliphatic. A total of just over 2 full repeats are present in PKD1 but they have been arranged into 3 incomplete repeats to show their similarity to those found in slit (Rothberg, 1990). The black boxes show identity to the LRR consensus and shaded boxes other regions of similarity between the repeats which have also been noted in other LRRs (Kobe, 1994).

Figure 16(b). The amino flanking region to the LRR in the PKD1 protein (33-71aa) is compared similar regions from a variety of other proteins. Black boxes shown identity with the consensus (adapted from [Rothberg, 1990 #1126]) and shaded boxes conserved amino acids. The different types of residue indicated in the consensus are: a, as above; p, polar or turn-like; h, hydrophobic. The listed proteins, with the species and Protein Identification Resource no. (PIR) shown in brackets, are: OMgp, oligodendrocyte myelin glycoprotein (Human, A34210); Slit (Drosophila; A36665); Chaoptin (Drosophila; A29943); GP-IB Beta, platelet glycoprotein 1bß chain (Human; A31929); Pg1, proteoglycan-1 (mouse; 520811); Biglycan (Human; A40757); Trk (Human; A25184) and LH-CF, lutropinchoriogonadotrophin receptor (Rat; A41343).

Figure 16(c). The carboxy flanking region of the LRR repeat from the PKD1 protein (126-180 aa) compared to similar regions in other proteins and a consensus accepted from [Rothberg, 1990 #1126]. The shading and amino acid types are as above. The proteins not described above are: Toll (Drosophila; A29943) and GP IX, platelet glycoprotein IX (Human; A46606).

Figure 17 is a sequence comparison of the C-type lectin domain. The PKD1 lectin domain (403-532aa) is compared to those of: BRA3, acorn barnacle lectin (JC1503); Kupffer cell carbohydrate-binding receptor (Rat; A28166), CSP, cartilage specific protoglycan (Bovine; A27752); Agp; asialoglycoprotein receptor (Human; 55283), E-Selectin (Mouse; B42755) and glycoprotein gp120 (Human; A46274). Black squares show identify with the consensus and shaded boxes conserved residues. Amino acid types are: Very highly conserved residues are shown in bold in the consensus which is adapted from Drickamer 1987, Drickamer 1988.

Figure 18 is a sequence analysis of the Ig-like repeat. The 16 copies of the PKD1 Ig-like repeat (PKDI 273-356 aa; PKDII-XVI, 851-2145aa) are compared to each other and to: V.a. colAi, and C.p. colA collagenases of Vibrio alginolyticus (S19658) and Clostridium perfringens (D13791), respectively; Pmel17, melanocyte specific glycoprotein (Human; A41234), FLT4, Ig repeat IV of fms-like tyrosine kinase 4 (Human; X68203), CaVPT, Ig repeat I of target protein of the calcium vector protein (CAVP) (amphioxius; P05548). black boxes shown amino acids identical in more

than 5 repeats and shaded boxes related residues. An Ig consensus determined from Harpaz et al. 1994 and Takagi et al. 1990 is shown in the symbols: a, aliphatic; h, hydrophobic; s, small and b, base with the predicted positions of the ß-strands indicated below. The PKD repeat IV has an extra repetition of 20 aa in the centre of the repeat while all of the others are between 84-87 aa.

Figure 19 reveals type III-related fibronectin domains. The four fibronectin-related domains from the PKD1 protein (2169-2573aa) are compared to similar domains in: Neuroglian (Drosophila; A32579); L1, neural recognition molecule L1 (X59847); F11, neural cell recognition molecule F11 (X14877); TAG 1, transiently expressed axonal surface glycoprotein-1 (Human; S28830); F3, Neuro-1 antigen (mouse; S05944); NCAM, neural cell adhesion molecule (Rat; X06564); DCC, deleted in colorectal cancer (Human; X76132); LAR, Leukocyte-common antigen related molecule (Human; Y00815); HPTP, ß protein tyrosine phosphate beta (Human; X54131) and FN, fibronectin (Human; X02761). The consensus sequence is compiled from Borh and Doolittle (1993), Kuma et al. (1993), Baron et al. (1992) and Borh and Doolittle (1992). Black boxes show identity to highly conserved residues and shaded boxes conserved changes or similarity in less highly conserved positions. The approximate positions of the ß strands are illustrated. The fibronectin repeats in the PKD1 protein are linked by sequences of 27aa (A-B), 22aa (B-C) and 7aa (C-D) which are not shown.

Figure 20 presents a proposed model of the PKD1

protein, polycystin. The predicted structure of the PKD1 protein is shown.

DETAILED DESCRIPTION

All references mentioned herein are listed in full at the end of the description which are herein incorporated by reference in their entirety. Except where the context clearly indicates otherwise, references to the PBP gene, transcript, sequence, protein or the like can be read as referring to the PKD1 gene, transcript, sequence, protein or the like, respectively.

A translocation associated with ADPKD

A major pointer to the identity of the PKD1 gene was provided by a Portuguese pedigree (family 77) with both ADPKD and TSC (Figure 1b). Cytogenetic analysis showed that the mother, 77-2, has a balanced translocation, 46XX t(16;22) (p13.3;q11.21) which was inherited by her daughter, 77-3. The son, 77-4, has the unbalanced karyotype, 45XY-16-22+der(16) (16qter-->16p13.3: :22q11.21-->2qter) and consequently is monosomic for 16p13.3-->16pter as well as for 22q11.21-->22pter. This individual has the clinical phenotype of TSC (see Experimental Procedures); the most likely explanation is that the TSC2 locus located within 16p13.3 is deleted in the unbalanced karyotype.

Further analysis revealed that the mother (77-2), and the daughter (77-3) with the balanced translocation, have the clinical features of ADPKD (see Experimental Procedures), while the parents of 77-2 were cytogenetically normal, with no clinical features of TSC and no renal cysts on ultrasound examination (aged 67 and 82 years). Although kidney cysts can be a feature of TSC, no other clinical

signs of TSC were identified in 77-2 or 77-3, making it unlikely that the polycystic kidneys were due to TSC. We therefore investigated the possibility that the translocation disrupted the PKD1 locus in 16p13.3 and proceeded to identify and clone the region containing the breakpoint.

The 77 family was analyzed with polymorphic markers from 16p13.3. Individual 77-4 was hemizygous for MS205.2 and GGG1, but heterozygous for SM6 and more proximal markers, locating the translocation breakpoint between GGG1 and SM6 (see Figure 1a). Fluorescence in situ hybridization (FISH) of a cosmid from the TSC2 region, CW9D (cosmid 1 in Figure 1a), to metaphase spreads showed that it hybridized to the der(22) chromosome of 77-2; placing the breakpoint proximal to CW9D and indicating that 77-4 was hemizygous for this region consistent with his TSC phenotype. DNA from members of the 77 family was digested with Cla I, separated by PFGE and hybridized with SM6; revealing a breakpoint fragment of about 100 kb in individuals with the der(16) chromosome (Figure 1c). The small size of this novel fragment enabled the breakpoint to be localized distal to SM6 in a region of just 60 kb (Figure 1a). A cosmid contig covering this region was therefore constructed (see Experimental Procedures for details).

The translocation breakpoint lies within a region duplicated elsewhere on chromosome 16p (16p13.1)

It is noted hereabove that the region between CW21 and N54 (Figure 1a) was duplicated at a more proximal site on

the short arm of chromosome 16 (Germino, et al., 1992; European Chromosome 16 Tuberous Sclerosis Consortium, 1993). Figure 2 shows that a cosmid, CW10III, from the duplicated region hybridized to two points on 16p; the distal, PKD1 region and a proximal site positioned in 16p13.1. The structure of the duplicated area is complex with each fragment present once in 16p13.3 re-iterated two-four times in 16p13.1 (see Figure 2). Cosmids spanning the duplicated area in 16p13.3 were subcloned (see Figure 3a and Experimental Procedures for details) and a restriction map was generated. A genomic map of the PKD1 region was constructed using a radiation hybrid, Hy145.19 which contains the distal portion of 16p but not the duplicate site in 16p13.1.

To localize the 77 translocation breakpoint, subclones from the target region were hybridized to 77-2 DNA, digested with Cla I and separated by PFGE. Once probes mapping across the breakpoint were identified they were hybridized to conventional Southern blots of 77 family DNA. Figure 3b shows that novel BamH I fragments were detected from the centromeric and telomeric side of the breakpoint, which was localized to the distal part of the probe 8S1 (Figure 3a). Hence, the balanced translocation was not associated with a substantial deletion, and the breakpoint was located more than 20 kb proximal to the TSC2 locus (Figure 3a). These results supported the hypothesis that polycystic kidney disease in individuals with the balanced translocation (77-2 and 77-3) was not due to disruption of the TSC2 gene, but

indicated that a separate gene mapping just proximal to TSC2, was likely to be the PKD1 gene.

The polycystic breakpoint (PBP) gene is disrupted by the translocation

Localization of the 77 breakpoint identified a precise region in which to look for a candidate or the PKD1 gene. During the search for the TSC2 gene we identified other transcripts not associated with TSC including a large transcript (about 14 kb) partially represented in the cDNAs 3A3 and AH4 which mapped to the genomic fragments CW23 and CW21 (Figure 3a). The orientation of the gene encoding this transcript had been determined by the identification of a polyA tract in the cDNA, AH4: the 3' end of this gene lies very close to the TSC gene, in a tail to tail orientation (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). To determine whether this gene crossed the translocation breakpoint genomic probes from within the duplicated area and flanking the breakpoint were hybridized to Northern blots. Probes from both sides of the breakpoint, between JH5 and JH13 identified the 14 kb transcript (Figure 3a and see below for details). Therefore, this gene, called 3A3, but not designated the PBP gene extended over the 77 breakpoint and consequently was a candidate for the PKD1 gene. A walk was initiated to increase the extent of the PBP cDNA contig and several new cDNAs were identified using probes from the single copy (non-duplicated) region (see Experimental Procedures for details). A cDNA contig was constructed which extended

about 5.7 kb, including about 2 kb into the area that is duplicated (Figure 3a).

Expression of the PBP gene

Initial studies of the expression pattern of the PBP gene were undertaken with cDNAs that map entirely within the single copy region (e.g. AH4 and 3A3). Figure 4a shows that the about 14 kb transcript was identified by 3A3 in various tissue-specific cell lines. From this and other Northern blots we concluded that the PBP gene was expressed in all of the cell lines tested, although often at a low level. The two cell lines which showed the highest level of expression were fibroblasts and a cell line derived from an astrocytoma, G-CCM. Significant levels of expression were also obtained in cell lines derived from kidney (G401) and liver (Hep3B). Measuring the expression of the PBP gene in tissue samples by Northern blotting proved difficult because such a large transcript is susceptible to minor RNA degradation. However, initial results with an RNAse protection assay, using a region of the gene located in the single copy area (see Experimental Procedures), showed a moderate level of expression of the PBP gene in tissue obtained from normal and polycystic kidney (data not shown). The widespread expression of the PBP is consistent with the systemic nature of ADPKD.

Identification of transcripts that are partially homologous to the PBP transcript

New cDNAs were identified with the genomic fragments, JH4 and JH8, that map to the duplicated region (Figure 3a

and see Experimental Procedures). However, when these cDNAs were hybridized to Northern blots a more complex pattern than that seen with 3A3 was observed. As well as the ~14 kb PBP transcript, three other, partially homologous transcripts were identified designated homologous gene-A (HG-A; ~21 kb), HG-B (~17 kb) and HG-C (8.5 kb) Figure 4b). There were two possible explanations for these results, either the HG transcripts were alternatively spliced forms of the PBP gene, or the HG transcripts were encoded by gene located in 16p13.1. To determine the genomic location of the HG loci a fragment from the 3' end of one HG cDNA (HG-4/1.1) was isolated. HG-4/1.1 hybridized to all three HG transcripts, but not to the PBP transcript and on a hybrid panel it mapped to 16p13.1 (not the PKD1 area). These results show that all the HG transcripts are related to each other outside the region of homology with the PBP transcript and that the HG loci map to the proximal site (16p13.1).

An abnormal transcript associated with the 77 translocation

As the PBP gene was transcribed across the region disrupted by the 77 translocation breakpoint, in a proximal to distal direction on the chromosome (see Figure 3a) it was possible that a novel transcript originating from the PBP promotor would be found in this family. Figure 4c shows that using a probe to the PBP transcript that mapped mainly proximal to the breakpoint, a novel transcript of approximately 9 kb (PBP-77) derived from the der(16) product of the translocation was detected. Interestingly, the PBP-77 transcript appears to be expressed at a higher level than

the normal PBP product. These results confirmed that the 77 translocation disrupts the PBP gene and supports the hypothesis that this is the PKD1 gene.

Mutations of the PBP gene in other ADPKD patients

To prove that the PBP gene is the defective gene at the PKD1 locus, we analyzed this region for mutations in patients with typical ADPKD. The 3' end of the PBP gene was most accessible to study as it maps outside the duplicated area. To screen this region BamH I digests of DNA from 282 apparently unrelated ADPKD patients were hybridized with the probe 1A1H.6, (see Figure 3a). In addition, a large EcoR I fragment (41 kb) which contains a significant proportion of the PBP gene was assayed by field inversion gel electrophoresis (FIGE) in 167 ADPKD patients, using the probe CW10. Two genomic rearrangements were identified in ADPKD patients by these procedures; each identified by both methods.

The first rearrangement was identified in patient OX875 (see Experimental Procedures for clinical details) who was shown to have a 5.5 kb genomic deletion without the 3' end of the PBP gene, producing a smaller transcript (PBP-875) (see Figures 5a, b and 3a for details). This genomic deletion results in a ~3 kb internal deletion of the transcript with the ~500 bp adjacent to the polyA tail intact. In this family linkage of ADPKD to chromosome 16 could not be proven because although OX875 has a positive family history of ADPKD there were no living, affected relatives. However, paraffin-embedded tissue from her

affected father (now deceased) was available. We demonstrated that this individual has the same rearrangement as OX875 by PCR amplification of a 220bp fragment spanning the deletion (data not shown). This result and analysis of two unaffected sibs of OX875, that did not have the deletion, showed that this mutation was transmitted with ADPKD.

The second rearrangement detected by hybridization was a 2 kb genomic deletion within the PBP gene, in ADPKD patient OX114 (see Experimental Procedures for clinical details and Figures 5c and 3a). No abnormal PBP transcript was identified by Northern blot analysis, but using primers flanking the deletion (see Experimental Procedures), a shortened product was detected by RT-PCR (Figure 5c). This was cloned and sequenced and shown to have a frame-shift deletion of 446 bp (between base pair 1746 and 2192 of the sequence shown in Figure 7). OX114 is the only member of the family with ADPKD (she has no children) and ultrasound analysis of her parents at age 78 (father) and 73 years old (mother) showed no evidence of renal cysts. Somatic cell hybrids were produced from OX114 and the deleted chromosome was found to be of paternal origin by haplotype analysis. The father of OX114 is now deceased but analysis of DNA from the brother of OX114 (OX984) with seven microsatellite markers from the PKD1 region (see Experimental Procedures) showed that he shares the same paternal chromosome, in the PKD1 region, as OX114. Renal ultrasound revealed no cysts in OX984 at age 53 and no deletion was detected by DNA

analysis (Figure 5c). Hence, the deletion in OX114 is a de novo event associated with the development of ADPKD. Although it is not possible to show that the ADPKD is chromosome 16-linked, the location of the PBP gene indicated that this is a de novo PKD1 mutation.

To identify more PKD1 associated mutations, single copy regions of the PBP gene were analyzed by RT-PCR using RNA isolated from lymphoblastoid cell lines established from ADPKD patients. cDNA from 48 unrelated patients was amplified with the primer pair 3A3 C (see Experimental Procedures) and the product of 260 bp was analyzed on an agarose gel. In one patient, OX32, an additional smaller product (125bp) was identified, consistent with a deletion or splicing mutation. OX32 comes from a large family in which the disease can be traced through three generations. Analysis of RNA from two affected sibs of OX32 and his parents showed that the abnormal transcript segregates with PKD1 (Figure 5d).

Amplification of normal genomic DNA with the 3A3 C primers generates a product of 418 bp; sequencing showed that this region contains two small introns (5', 75 bp and 3', 83 bp) flanking a 135 bp exon. The product amplified from OX32 genomic DNA was normal in size, excluding a genomic deletion. However, heteroduplex analysis of that DNA revealed larger heteroduplex bands, consistent with a mutation within that genomic interval. The abnormal OX32, RT-PCR product was cloned and sequenced: this demonstrated that, although present in genomic DNA, the 135 bp exon was

missing from the abnormal transcript. Sequencing of OX32 genomic DNA demonstrated a G-->C transition at +1 of the splice donor site following the 135 bp exon. This mutation was confirmed in all available affected family members by digesting amplified genomic DNA with the enzyme Bst NI: a site is destroyed by the base substitution. The splicing defect results in an in-frame deletion of 135 bp from the PBP transcript (3696 bp to 3831 bp of the sequence shown in Figure 7). Together, the three intragenic mutations confirm that the PBP gene is the defective gene at the PKD1 locus.

Deletions that disrupt the TSC2 and the PKD1 gene

The deletion called WS-53 disrupts both the TSC2 gene and the PKD1 gene (European Chromosome 16 Tuberous Sclerosis Consortium, 1993), although the full proximal extent of the deletion was not determined. Further study has shown that the deletion extends ~100 kb (see Figure 6 for details) and deletes most if not all of the PKD1 gene. This patient has TSC but also has unusually severe polycystic disease of the kidneys. Other patients with a similar phenotype have also been under investigation. Deletions involving both TSC2 and PKD1 were identified and characterized in six patients in whom TSC was associated with infantile polycystic kidney disease. As well as the deletion in WS-53, those in WS-215 and "S-250 also extended proximally well beyond the known distribution of PKD1 and probably delete the entire gene. The deletion in WS-194 extended over the known extent of PKD1, but not much further proximally, while the proximal breakpoints in WS-219 and WS-227 lay within PKD1 itself.

Northern analysis of case WS-219 with probe JH8, which lies outside the deletion, showed a reduced level of the PKD1 transcript but no evidence of an abnormally sized transcript (data not shown). Analysis of samples from the clinically unaffected parents of patients WS-53, WS-215, WS-219, WS-227 and WS-250 showed the deletions in these patients to be de novo. The father of WS-194 was unavailable for study.

In a further case (WS-212), renal ultrasound showed no cysts at four years of age but a deletion was identified which removed the entire TSC2 gene and deleted an XbaI site which is located 42 bp 5' to the polyadenylation signal of PKD1. To determine the precise position of the proximal breakpoint in PKD1, a 587bp probe from the 3' untranslated region (3'UTR) was hybridized to XbaI digested DNA. A 15kb XbaI 1 breakpoint fragment was detected with an approximately equal intensity to the normal fragment of 6kb, indicating that most of the PKD1 3'UTR was preserved on the mutant chromosome. Evidence that a PKD1 transcript is produced from the deleted chromosome in WS-212 was obtained by 3' rapid identification of cDNA ends (RACE) with a novel, smaller product generated from WS-212 cDNA. Characterization of this product showed that polyadenylation occurs 546bp 5' to the normal position, within the 3'UTR of PKD1 (231bp 3' to the stop codon at 5073bp of the described PKD1 sequence[14]). A transcript with an intact open reading frame is thus produced from the deleted WS-212 chromosome. It is likely that a functional PKD1 protein in produced from this transcript, explaining the lack of cystic disease in

this patient. The sequence preceding the novel site of polyA addition is:

AGTCAGTAATTTATATGGTGTTAAAATGTG(A)n.

Although not conforming precisely to the consensus of AATAAA, it is likely that part of this AT rich region acts as an alternative polyadenylation signal if, as in this case, the normal signal is deleted (a possible sequence is underlined).

The WS-212 deletion is 75kb between SM9-CW9 distally and the PKD1 3'UTR proximally. The WS-215 deletion is 160kb between CW15 and SM6-JH17. WS-194 has 65kb deleted between CW20 and CW10-CW36. WS-227 has a 50kb deletion between CW20 and JH11 and WS-219 has a 27kb deletion between JH1 and JH6. The distal end of the WS-250 deletion is in CW20 but the precise location of the proximal end is not known. However, the same breakpoint fragment of 320kb is seen with PvuI-digested DNA using probes on adjacent PvuI fragments, CE18 (which normally detects a 245kb fragment) and Blu24 (235kb). Hence this deletion can be estimated ~160kb. b. PFGE analysis of the deletion in WS-219. MluI digested DNA from a normal control (N) and WS-219 probed with the clones H2, JH1, CW21 and CW10 which detect an ~130kb fragment in normal individuals. CW10 also detects a much smaller fragment from the duplicated region situated more proximally on 16p. A novel fragment of ~100kb is seen in WS-219 with probes H2 and CW10 which flank the deletion in this patient. JH1 is partially deleted but detects the novel band weakly. The aberrant fragment is not detected by CW-21, which is deleted

on the mutant chromosome. BamHl digested DNA of normal control (N) and WS-219 separated by conventional gel electrophoresis and hybridized to probes JH1 and JH6 which flank the deletion. The same breakpoint fragment of ~3kb is seen with both probes, consistent with a deletion of ~27kb ending within the BamHl fragments seen by these probes.

Two further deletions

In addition we have characterized two further mutations of this gene which were identified in typical PKD1 families. In both cases the mutation is a deletion in the 75bp intron amplified by the primer pair 3A3C (European Polycystic Kidney Disease Consortium, 1994). The deletions are of 18bp and 20bp, respectively, in the patients 461 and OX1054. Although these deletions do not disrupt the highly conserved sequences flanking the exon/intron boundaries, they do result in aberrant splicing of the transcript. In both cases, two abnormal mRNAs are produced, one larger and one smaller than normal. Sequencing of these cDNAs showed that the larger transcript includes the deleted intron, and so has an in-frame insertion of 57bp in 461, while OX1054 has a frameshift insertion of 55bp. The smaller transcript is due to activation of a cryptic splice site in the exon preceding the deleted intron and results in an in-frame deletion of 66bp in both patients. The demonstration of two additional mutations of this gene in PKD1 patients further confirms that this is the PKD1 gene.

Partial Characterization of the PKD1 gene

To characterize the PKD1 gene further, evolutionary

conservation was analyzed by 'zoo blotting'. Using probes from the single copy, 3' region (3A3) and from the duplicated area (JH4, JH8) the PKD1 gene was conserved in other mammalian species, including horse, dog, pig and rodents (data not shown). No evidence of related sequences were seen in chicken, frog or drosophila by hybridization at normal stringency. The degree of conservation was similar when probes from the single copy of the duplicated region were employed.

Although the full genomic extent of the PKD1 gene was not yet known, results obtained by hybridization to Northern blots showed that it extended from at least as far as JH13. Several CpG islands were localized 5' of the known extent of the PKD1 gene (Figure 6), although there was no direct evidence that any of these are associated with this gene.

The cDNA contig extending 5631 bp to the 3' end of the PKD1 transcript was sequenced; where possible more than one cDNA was analyzed and in all regions both strands were sequenced (Figure 7). We estimated that this accounts for ~40% of the PKD1 transcript. An open reading frame was detected which runs from the 5' end of the region sequenced and spans 4842 bp, leaving a 3' untranslated region of 789 bp which contains the previously described microsatellite, KG8 (Peral, et al., 1994; Snarey, et al., 1994). A polyadenylation signal is present at nucleotides 5598-5603 and a polyA tail was detected in two independent cDNAs (AH4 and AH6) at position, 5620. Comparison with the cDNAs HG-4 and 11BHS21, which are encoded by genes in the duplicate,

16p13.1 region, show that 1866 bp at the 5' end of the partial PKD1 sequence shown in Figure 7 lies within the duplicated area. The predicted amino acid sequence from the available open reading frame extends 1614 residues, and is shown in Figure 7. A search of the swissprot and NBRF data bases with the available protein sequence, using the Blast program (Altschul, et al., 1990) identified only short regions of similarity (notably, between amino-acids 690-770 and 1390-1530) to a diverse group of proteins; no highly significant areas of homology were recognized. The importance of the short regions of similarity is unclear as the search for protein motifs with the ProSite Program did not identify any recognized functional protein domains within the PKD1 gene.

The test of identifying and characterizing the PKD1 gene has been more difficult than for other disorders because more than three quarters of the gene is embedded in a region of DNA that is duplicated elsewhere on chromosome 16. This segment of 40-50 kb of DNA, present as a single copy in the PKD1 area (16p13.3), is re-iterated as several divergent copies in the more proximal region, 16p13.1. This proximal site contains three gene loci (HG-A, -B and -C) that each produce polyadenylated mRNAs and share substantial homology to the PKD1 gene; it is not known whether these partially homologous transcripts are translated into functional proteins.

Although gene amplification is known as a major mechanism for creating protein diversity during evolution,

the discovery of a human disease locus embedded within an area duplicated relatively recently is a new observation. In this case because of the recent nature of the reiteration the whole duplicated genomic region retains a high level of homology, not just the exons. The sequence of events leading to the duplication and which sequence represents the original gene locus are not yet clear. However, early evidence of homology of the 3' ends of the three HG transcripts which are different from the 3' end of the PKD1 gene indicated that the loci in 16p13.1 have probably arisen by further reiteration of sequences at this site, after it separated from the distal locus.

To try to overcome the duplication problem we employed an exon linking approach using RNA isolated from a radiation hybrid, HY145.19, that contains just the PKD1 part of chromosome 16, and not the duplicate site in 16p13.1. Hence, this hybrid produces transcripts from the PKD1 gene but not from the homologous genes (HG-A, HG-B and HG-C). We have also sequenced much of the genomic region containing the PKD1 gene, from the cosmid JH2A, and have sequenced a number of cDNAs from the HG locus. To determine the likely position of PKD1 exons in the genomic DNA we compared HG cDNAs, (HG-4 and HG-7) to the genomic sequence. We then designed primers with sequences corresponding to the genomic DNA, to regions identified by the HG exons and employing DNA generated from the hybrid HY145.19, we amplified sections of the PKD1 transcript. The polymerase Pfu was used to minimise incorporation errors. These amplified fragments

were then cloned and sequenced. The PKD1 cDNA contig whose sequence is shown in Figure 10 is made up of (3'-5') the original 5.7 kb of sequence shown in Figure 7, and the cDNAs: gap α 22 (890 bp), gap gamma (872 bp), a section of genomic DNA from the clone JH8 (2,724 bp) which corresponds to a large exon, S1-S3 (733 bp), S3-S4 (1,589 bp) and S4-S13 (1,372 bp). Together these make a cDNA of 13,807nt. When these cDNAs from the PKD1 contig were sequenced an open reading frame was found to run from the start of the contig to the stop codon, a region of 13,018 bp. The predicted protein encoded by the PKD1 transcript is also shown in Figure 10 and has 4,339 amino acid residues.

Cloning a full length PKD1 cDNA

cDNAs known to originate from the PKD1 or HG transcripts show on average a sequence divergence of less than 3%. Consequently, although many cDNAs were identified by hybridisation of various PKD1 genomic probes to cDNA libraries, it proved difficult to differentiate genuine PKD1 clones from those of the HG transcripts. For this reason a novel strategy was employed to clone the PKD1 transcript.

To obtain a template of genomic sequence of the PKD1 gene, clones which contain the transcribed region, JH6 and JH8-JH13, were sequentially truncated and sequenced. These clones were isolated from the cosmid JH2A, which extends into the single copy area containing the 3' portion of the PKD1 gene (figure 13) and hence represents the PKD1 and not the HG loci. As a result of this analysis a contig of about 18 kb of genomic sequence was generated, which was

ultimately found to encode >95% of the unsequenced portion of the PKD1 transcript.

A number of HG cDNA clones identified by the DNA probes JH8 or JH13 (including HG-4, HG-7C and 13A1) were sequenced. Clones identified by JH8 were chosen because this genomic area is duplicated fewer times than the surrounding DNA, with only the HG-A and HG-B transcripts (not HG-C) homologous to this region. The comparison of these cDNA and genomic sequences showed a characteristic intron/exon pattern and we concluded that the exons highlighted in the genomic sequence were likely to be exons of the PKD1 gene. To prove this, pairs of primers matching the sequence of the putative PKD1 exons and spaced 0.7 - 2 kb apart in the proposed transcript, were synthesised. Employing RNA from a radiation hybrid, HY145.19, that contains the PKD1 but not the HG loci, PKD1 specific cDNAs were amplified by RT-PCR and cloned (see Experimental Procedures for details). In this way, a number of overlapping cDNAs spanning the PKD1 transcript, for the cDNAs at the 3' end to those homologous to JH13 were cloned (Figure 13).

Analysis of a further cDNA, HG-6 showed that a short region (~100 bp) of HG-6 lay 5' to the sequenced genomic region and this was located by hybridisation to the genomic clone SM3 (figure 13); SM3 was subsequently sequenced. The position of the cDNA in SM3 was identified and the possible 5' extent of this exon was determined in the genomic sequence; and in-frame stop codon was identified hear the 3' end of the exon. This exon lay at a CpG island (described

hereinafter) suggesting, along with the presence of the stop codon, that this may be the first exon of the PKD1 gene. to determine the likely transcriptional start site the method of primer extension from three different oligos within the first exon was employed (see Experimental Procedures). In all cases, a transcriptional start was identified at the same G nucleotide and showed the first exon to be 426 bp. The structure of the PKD1 transcript was confirmed by a final exon link, rev1 which starts 3 bp 3' to the proposed transcriptional start (see figure 13 and Experimental Procedures for details).

The intron/exon structure of the PKD1 gene

Sequencing the cDNA contig revealed a total sequence of 14, 148 bp which extends over approximately 52 bp of genomic sequence from SM3 to BFS5 (Figure 13). We were able to determine the intron/exon structure of much of the gene by direct comparison between the cDNA and genomic sequence. In the 3' region of the gene (JH5-BFS5), a partial genomic sequence was obtained at intron/exon borders by sequencing the corresponding genomic clone from exonic primer.

The PKD1 CpG island

The 5' end of the gene lies at CpG island SM3. SM3 is located entirely within the duplicated region, but this clone was isolated from the cosmid SM11 which extends through the duplicated area into the proximal flanking single copy region and therefore is known to originate from this area. Figure 14 shows a map of the PKD1 CpG island including genomic sites for several methylation sensitive

enzymes, the location of the first exon and the GC content across the island. Evidence that the enzyme sites in the PKD1 region (and not just the HG area) digest, was obtained by pulsed field gel electrophoresis with the enzymes Mlu I, Not II and BssH II using probes outside the duplicated area. Digestion of the Sac II sites and confirmation of the Not I site was made with a panel of somatic cell hybrids which either contain just the HG (P-MWH2A) or just the PKD1 locus (Hy145.19). These results showed that the Sac II and Not I sites digest in both sets of hybrids (data not shown), indicating that this region is a CpG island in the HG as well as the PKD1 area. Further proof that this is the likely position of a functional promoter was obtained by analysis for DNAase 1 hypersensitivity. A DNAase hypersensitive site in the region 5' to the transcription start site in SM3 was detected (figures 14a and b).

Analysis of the PKD1 transcript

Analysis of the sequence shows an open reading frame running from the start of the sequence to position 13,117 bp (Figure 15). Detailed sequencing of the genomic region containing the 3' portion of the gene revealed two extra Cs at positions 13,081-2 (Figure 15). An in-frame start codon which is consistent with the Kozak consensus was detected at position 212 bp; just 3' to the stop codon in the 5'UTR. Analysis for a signal sequence cleavage site using the von Hinge (von Hinge 1986) algorithm showed a high probability of a hydrophobic signal sequence with cleavage at amino acid 23 (see Figure 15). The total length of the predicted

protein is 4302 aa with a calculated molecular mass after excision of the signal peptide of 460 kD and an estimated isoelectric point of 6.26. However, this may be an underestimate of the total mass of the protein as many potential sites for N-linked glycosylation are present (Figure 15).

Homologies with the PKD1 protein

The predicted PKD1 protein was analysed for homologies with know proteins in the SwissProt and NBRF databases using the BLAST Altschul et al 1990) and FASTA algorithms. This analysis revealed two clear homologies and also a number of other potential similarities which were studied on detail.

Leucine rich repeat

Near the 5' end of the PKD1 protein is a region of leucine rich-repeats (LRRs). LRRs are a highly conserved motif usually of 24 residues with precisely spaced leucines (or other aliphatic amino acids) and an asparagine at position 19 (Figure 16a and reviewed in Kobe and Reisenhofer (1994)). Two complete LRRs plus a partial repeat unit are found in the PKD1 protein, which have complete homology with the LRR consensus.

Surrounding the LRRs are distinctive cysteine-rich amino and carboxy flanking regions (Figures 16b and c). This flank-LRR-flank structure is exclusively found on proteins in extracellular locations and is thought to be involved in protein-protein interactions such as adhesion to other cells or to components of the extracellular matrix or as a receptor concerned with binding or signal transduction.

The structure found in the PKD1 protein is similar to that found in the Drosophila protein, slit, which is important for normal central nervous system development (Rothberg, 1990). Although slit contains far more LRRs than the PKD1 protein, with four blocks each consisting of 4 or 5 repeat units, the structure of each block is similar as they finish on the amino and carboxy side with shortened LRRs which are immediately flanked by the cysteine rich regions. In the PKD1 protein two shortened LRRs surround one complete repeat unit and immediately abut the amino and carboxy flanking regions.

The amino flanking region consists of four invariant cysteines and a number of other highly conserved residues in an area of 30-40 amino acids; comparison of the PKD1 region to amino flanking motifs of other proteins is shown in figure 4b. The carboxy flanking region extends over an area of between 50-60 residues and consists of an invariant proline and four cysteines plus several other highly conserved amino acids. The similarity of the PKD1 region to carboxy flanking regions from other proteins is shown in figure 4c.

Some LRR proteins, such as slit (Rothberg 1990) and small proteoglycans are wholly extracellular but others including Toll (Hashimoto et al, 1990) and trkc (Lamballe 1991) have a single transmembrane sequence, while the LH-CRG receptor and related proteins have seven trans-membrane segments and are involved in signal transduction.

C type lectin domain

Analysis of the sequence from exons 6 and 7 showed a high level of homology with a C type lectin domain. C-type lectins are found in a variety of proteins in extracellular locations where they bind specific carbohydrates in the presence of Ca^{2+} ion (Drickamer 1987, 1988; Weiss 1992). Figure 17 illustrates the similarity of the PKD1 lectin domain to those found in a number of proteins including: proteogylcans, which interact with collagens and other components of the extracellular matrix; endocytic receptors, and selectins which are involved in cell adhesion and recognition. Three different selectins have been identified: E-selectin (endothelium), P-selectin (platelets) and L-selectin (lymphocytes) and these work with other cell adhesion molecules to promote binding of the cell carrying the selectin to various other target cells.

Immunoglobulin-like repeat motif

Significant homologies were detected between a region of exon 5 and three regions of exon 15, with the same conserved sequence, WDFGDGS, which is also found in a melanocyte-specific secreted glycoprotein, Pmel17 (Kwon et al, 1991) and three prokaryotic collagenases or proteinases (Ohara et al, 1989, Takeuchi et al, 1992 and Matsushita et al, 1994). Further analysis of the amino acid sequence of the PKD1 protein showed that a conserved region of approximately 85 bp could be discerned around this central sequence and that 16 copies of this repeat were present in the PKD1 protein; 1 in exon 5 and the other 15 as a tandem array in exons 11 to 15. Figure 18 shows that a highly

conserved structure is maintained between the repeats although in some cases less similarity is noted with the WDFGDGS sequence. Further analysis of the most conserved residues found in the repeat units showed similarity to various immunoglobulin (Ig) domains; two Ig repeats which show particular homology to the PKD1 protein are shown (figure 18). The repeat unit is most similar to that found in a number of cell adhesion and surface receptors which have recently been defined as the I set of Ig domains (Harpaz 1994). Ig repeats consist of 7-9 β strands of 5-10 residues linked by turns which are packed into two β sheets. The B, C, F and G β-strands of the I set are particularly similar to the PKD1 repeat, although the highly conserved cystine residues which stabilise the two β sheets through a disulphide bond are absent. The D and E β strands, however, seem less similar and in some cases are significantly shortened or apparently absent. Further evidence that this PKD1 repeat has an Ig-like structure is found by analysis of the secondary structure with the predominant configuration found of β strands linked by turns. The WDFGDS area of the Ig molecule is one that often has a specific binding function (Jones et al., 1995) and this sequence may have a specific binding role in polycystin.

Type III fibronectin-related domains

Analysis of the secondary structure of the PKD1 protein beyond the carboxy end of the region of Ig-like repeats showed a continuation of the β stand and turn structure. No evidence of further Ig-like repeats could be found in this

area but three pairs of evenly spaced (38-40aa) tryptophan and tyrosine residues was noted which are the most highly conserved positions of the type III fibronectin repeat which has a similar secondary structure to Ig domains. Further analysis and comparison with other type III fibronectin domains showed that in total four fibronectin repeats (one with leucine replacing the conserved tyrosine) could be recognised in this area with many of the most highly conserved residues of this domain found in the PKD1 repeat (Figure 20).

A large number of proteins with Ig-like repeats have now been described which are involved in cell-cell interactions and cell adhesion (reviewed in Brummendork and Rathjen, 1994), while type III fibronectin (FNIII) domains are found on extracellular matrix molecules and adhesion proteins. A number of cell adhesion proteins which are located mainly on neural cells, have both Ig-like and FNIII-related domains. In these cases the FNIII repeats are always positioned C-terminal of the Ig-like units and close to a transmembrane domain; a similar pattern is seen in the proposed structure of polycystin. These Ig/FNIII containing proteins such as neuroglican and NrCAM are thought to be involved in neuron-neuron interactions and the patterning of the axonal network.

Many cell adhesion proteins of the Ig superfamily are also involved in communication and signal transduction mediated through their cytoplasmic tails. These cytoplasmic regions are known to bind to cytoskeletal proteins and other

intracellular components, and phosphorylation of this part of the molecule is also thought to affect adhesive properties of the protein; potential phosphorylation sites are found in the cytoplasmic tail and one intracellular loop of polycystin (Figure 20).

Transmembrane regions

Analysis of hydrophobicity predicted that the deduced protein is an integral membrane protein with a signal peotide and multiple transmembrane (TM) domains located in the C-terminal region. From this analysis 11 regions (including the signal peptide) had a mean hydrophobicity indice higher than 1.4 and therefore were considered as certain membrane spanning domains (see Experimental Procedures for details). Three others with a mean hydrophobicity indice between 0.75-1.0 were considered as putative TM domains. The most likely topology of the protein was predicted using TopPed II programme (see Experimental Procedures for details) and the resulting model included one putative segment plus the 10 certain transmembrane domains and the signal peptide. According to this model the N-terminal end is extracellular and the (highly hydrophobic) carboxy-terminal region is anchored to the membrane by 11 membrane-spanning segments, with the highly charged carboxy end located in the cytoplasm. This topology is supported by the study of N-glycosylation sites with all but one site, out of a total of 61 predicted, in an extracellular location according to the model, including 11 in the two large extracellular loops between TM regions.

However, if degree of hydrophobicity required to define a certain putative transmembrane region is altered within the model, the predicted number of such domains can change to 9 (excluding the most N-terminal pair) or 13 (with two new domains defined between TM7 and TM8). This can be ascertained by studies with specific antibodies.

Most transmembrane proteins containing the types of cell adhesion domain found on polycystin have a single transmembrane domain. The role of the multiple membrane spanning domains found in polycystin is not yet clear.

Proposed structure of the PKD1 protein

From the detailed analysis of the predicted PKD1 protein sequence a model of the likely structure of the protein can be formulated (Figure 20). This model predicts an extracellular N-terminal region of approximately 2550 aa containing several distinctive extracellular domains and an intracellular C-terminus of approximately 225 aa. The intervening region of nearly 1500 aa is associated with the membrane with 11 transmembrane regions predicted and 10 variously sized extracellular and cytoplasmic loops (see Figure 20). A proline rich hinge is found between the flank-LRR-flank region and the first Ig-like repeat. Two phosphorylation sites for tyrosine kinase and protein kinase C are found in cytoplasmic locations (Figures 15 and 20).

Therefore, the PKD1 protein, named polycystin, has highlighted several clear domains, plus a reiterated motif that occupies over 30% of the protein.

Characterisation of the PKD1 gene has proven to be a

uniquely difficult problem because most of the gene lies in a region which is reiterated elsewhere on the chromosome. The high degree of similarity between the two areas ()97%) both in exons and introns has meant that a novel approach has been required to clone the full length transcript; involving extensive genomic sequencing and generating cDNAs from a cell line with the PKD1 but not the HG loci. In this way a contig containing the entire PKD1 transcript has now been cloned.

Preliminary analysis shows that the HG genes are very similar to PKD1 both in terms of genomic structure and sequence over most of their length (apart from the novel 3' regions). The 5' end of the PKD1 gene is at a CpG island which lies within the duplicated area. Homologous areas to this island, in the HG region, also have cleavable sites for methylation sensitive enzymes; these duplicate islands probably lie at the 5' ends of the various HG genes. Analysis for DNAase hypersensitivity also indicates that the HG, CpG islands probably contain active promoters. These results are consistent with the observation of polyadenylated mRNA from the HG genes on Northern blots and the similarity of the expression pattern of the HG and PKD1 genes in different tissue specific cell lines. The HG genes may have complete open reading frames and may encode functional proteins. Antibodies to their 'unique' 3' regions will be required to determine this. Although the PKD1 transcript is large, the overall size of the gene, at 52 kb, is not (the Duchenne muscular dystrophy (DMD) gene which

encodes a slightly smaller transcript has a genomic size of over 2Mb). Indeed, if the first intron of PKD1 is excluded from the analysis, 40.3% of the remainder of the gene is found in the mature mRNA. In the compact structure of the PKD1 gene, some of the introns are close to or smaller than the minimal size of 80 bp thought to be required for efficient splicing, although they are presumably excised effectively. We have shown that deletion of 18 or 10 bp from one small intron (intron 43), resulting in an intron of 55 or 57 bp, leads to aberrant splicing (Peral, 1995). Similar mutations may be found in the other small introns of this gene. The compact nature of the PKD1 gene probably reflects the GC rich area of the genome in which it is found (the PKD1 transcript has a total GC content of about 65%); a similar organisation is seen in other genes from the area of chromosome 16 (Vyas, 1992) is in an AT rich genomic region.

It is clear that polycystin has many features of a cell adhesion or recognition molecule with multiple different extracellular domains. These various binding domains are likely to have different specificities so that it can be envisaged that it will bind to a variety of different proteins (and carbohydrates) both on other cells and possibly in the extracellular matrix. Although provisional evidence indicates a wide range of expression of polycystin in tissue specific cell lines. detailed analysis by in situ of the mRNA and with antibodies to determine the cells expressing this protein both in adult tissue and during

development will provide further evidence.

Initial analysis has revealed little clear evidence of alternate splicing; although one cDNA (out of 6 studied) had an extra exon of 255 bp positioned in intron 16. This exon contains an in-frame stop codon and it is not known at this stage if this represents an incompletely spliced mRNA or a splice form of polycystin which terminates at this point. Truncation of the protein here would leave a secreted protein lacking all of the transmembrane and cytoplasmic regions. Interestingly, a similar secreted form of the neural adhesion protein, NCAM, which is normally attached to the cell membrane, is produced by alternate splicing by insertion of an exon containing a stop codon (Gower et al., 1988).

The initial changes that have been noted in ADPKD kidneys are abnormal thickening and splitting of the basement membrane (BM) and simultaneous de-differentiation of associated epithelial cells at the point of tubular dilation. Similar results have been noted in the heterozygote Han:SPRD rat (Schafer et al., 1994) which is a dominant model of PKD, although it is not known if it is a rat model of PKD1. Concurrent changes in cellular characteristics and the BM suggests that a disruption or alteration of communication between the cell and the BM may be the primary change in this disease. Polycystin could play an important role in interaction and communication between epithelial cells and the BM. It is known that signals are required from cells to the extracellular matrix

(ECM) for normal BM development and also that communication from the ECM to cells is required for control of cellular differentiation. Communication between the ECM and cells occurs by several different means including through integrins and so polycystin may bind to integrins, although it may interact directly with components of the ECM. Although ADPKD is generally a disease of adulthood, there is plenty of evidence that the cystic changes in the kidney may start much earlier (Milutinovic et al., 1970), even in utero (Reeders, 1986). Expression of polycystin during renal development may be when its major role occurs, perhaps in assembly of the BM and it is then that the errors, which later lead to cyst development, occur.

The plethora of connective tissue abnormalities associated with ADPKD indicate that the adhesion/communication roles of polycystin may be important for assembly and/or maintenance of the BM in many tissues, as well as the kidney. Hence, it is possible that disruption of normal cell adhesion and communication mediated by polycystin may explain the primary defects seen in the kidney and other organs in ADPKD. Clearly molecules that interact with polycystin or have a similar role are candidates for the other renal polycystic diseases of man.

A study of the mutations of the PKD1 gene highlight important functional regions of the protein. All of the mutations described so far in typical PKD1 families involve deletion or other disruption in the 3' end of gene. Two large deletions detected on Southern blots remove a large

part of the protein (or make an out of frame product) including the last 6 transmembrane domains and the C-terminal end. The in-frame splicing change described in the same paper would remove most of TM10 and part of the preceding cytoplasmic loop. Two recently described splicing mutations (Peral, 1995) create three different products which either delete part of the cytoplasmic loop between TM7 and TM8 or a larger region of this loop including part of TM7 or insert an extra region into that loop. These mutated genes may make functional protein (they all produce abnormal mRNA) and it is interesting to note that, in each case, these proteins would have an intact extracellular region with disrupted cytoplasmic and transmembrane areas. Such proteins may bind to extracellular targets but are unable to communicate in a normal way.

A group of mutations of PKD1 which completely delete the gene and hence are clearly inactivating have been described (Brook-Carter, 1994). However, in each of these cases the deletions also disrupt the adjacent TSC2 gene making interpretation of these cases difficult (TSC2 mutations alone can cause the development of renal cysts). Nevertheless, the severity of the polycystic disease in these patients indicate that inactivation of one PKD1 allele does promote cyst development. Further more, all these children are often severely affected at birth, cyst formation must occur in utero in these cases and hence polycystin has an important developmental role. A second somatic hit in the target tissue may also be required in

these cases (and normal PKD1 patients) before cyst development can occur.

PKD1 GENE AND POLYCYSTIC KIDNEY DISEASE

We have therefore compelling evidence that mutations of the PKD1 gene give rise to the typical phenotype of ADPKD. The location of this gene within the PKD1 candidate region and the available genetic evidence from the families with mutations show that this is the PKD1 gene. The present invention therefore includes the complete PKD1 gene itself and the six PKD1 - associated mutations which have been described: a de novo translocation, which was subsequently transmitted with the phenotype; two intragenic deletions (one a de novo event); two further deletions; and a splicing defect.

It has been argued that PKD1 could be recessive at the cellular level, with a second somatic mutation required to give rise to cystic epithelium (Reeders, 1992). This "two hit" process is thought to be the mutational mechanism giving rise to several dominant diseases, such as neurofibromatosis (Legius, et al., 1993) and tuberous sclerosis (Green, et al., 1994) which result from a defect in the control of cellular growth. If this were the case, however, we might expect that a proportion of constitutional PKD1 mutations would be inactivating deletions as seen in these other disorders.

The location of the PKD1 mutations may, however, reflect some ascertainment bias as it is this single copy area which has been screened most intensively for mutations.

Nevertheless, no additional deletions were detected when a large part of the gene was screened by FIGE, and studies by PFGE showed no large deletions of this area in 75 PKD1 patients. It is possible that the mutations detected so far result in the production of an abnormal protein which causes disease through a gain of function. However, it is also possible that these mutations eliminate the production of functional protein from this chromosome and result in the PKD1 phenotype by haploinsufficiency, or only after loss of the second PKD1 homologue by somatic mutation.

At least one mutation which seems to delete the entire PKD1 gene has been identified (WS-53) but in this case it also disrupts the adjacent TSC2 gene and the resulting phenotype is of TSC with severe cystic kidney disease. Renal cysts are common in TSC so that the phenotypic significance of deletion of the PKD1 gene in this case is difficult to assess. It is clear that not all cases of renal cystic disease in TSC are due to disruption of the PKD1 gene; chromosome 9 linked TSC (TSC1) families also manifest cystic kidneys and we have analysed many TSC2 patients with kidney cysts who do not have deletion of the PKD1 gene.

Preliminary analysis of the PKD1 protein sequence has highlighted two regions which provide some clues to the possible function of the PKD1 gene. At the extreme 5' end of the characterised region are two leucine-rich repeats (LRRs) (amino acids 29-74) flanked by characteristic amino flanking (amino acids 6-28) and carboxy flanking sequences

(amino acids 76-133) (Rothberg et al., 1990). LRRs are thought to be involved in protein-protein interations (Kobe and Deisenhofer, 1994) and the flanking sequences are only found in extracellular proteins. Other proteins with LRRs flanked on the amino and carboxy sides are receptors or are involved in adhesion or cellular signalling. Further 3' on the protein (amino acids 350-515) is a C-type lectin domain (Curtis et al., 1992). This indicates that this region binds carbohydrates and is also likely to be extracellular. These two regions of homology indicate that the 5' part of the PKD1 protein is extracellular and involved in protein-protein interactions. It is possible that this protein is a constituent of, or plays a role in assembling, the extracellular matrix (ECM) and may act as an adhesive protein in the ECM. It is also possible that the extracellular portion of this protein is important in signalling to other cells. The function of much of the PKD1 protein is still not fully known but the presence of several hydrophobic regions indicates that the protein may be threaded through the cell membrane.

Familial studies indicate that de novo mutations probably account for only a small minority of all ADPKD cases; a recent study detected 5 possible new mutations in 209 families (Davies, et al., 1991). However in our study one of three intragenic muttions detected was a new mutation and the PKD1 associated translocation was also a de novo event. Furthermore, the mutations detected in the two familial cases do not account for a significant proportion

of the local PKD1. The OX875 deletion was only detected in 1 of 282 unrelated cases, and the splicing defect was seen in only 1 of 48 unrelated cases. Nevertheless, studies of linkage disequilibrium have found evidence of common haplotypes associated with PKD1 in a proportion of some populations (Peral, et al., 1994; Snarey, et al., 1994) suggesting that common mutations will be identified.

Once a larger range of mutations have been characterised it will be possible to evaluate whether the type and location of mutation determines disease severity, and if there is a correlation between mutation and extra-renal manifestations. Previous studies have provided some evidence that the risk of cerebral aneurysms 'runs true' in families (Huston, et al., 1993) and that some PKD1 families exhibit a consistently mild phenotype (Ryynanen, et al., 1987). A recent study has concluded that there is evidence of anticipation in ADPKD families, especially if the disease is transmitted through the mother (Fink, et al., 1994). Furthermore, analysis of families with early manifestations of ADPKD show that there is a significant intra-familial recurrence risk and that childhood cases are most often transmitted maternally (Rink, et al., 1993; Zerres, et al., 1993). This pattern of inheritence is reminiscent of that seen in diseases in which an expanded trinucleotide repeat was found to be the mutational mechanism (reviewed in Mandel, 1993). However, no evidence for an expanding repeat correlating with PKD1 has been found in this region although such a sequence cannot be excluded.

There is ample evidence that early presymptomatic diagnosis of PKD1 is helpful because it allows complications such as hypertension and urinary tract infections to be monitored and treated quickly (Ravine, et al., 1991). The identification of mutations within a family allow rapid screening of that and other families with the same mutation. However, genetic linkage analysis is likely to remain important for presymptomatic diagnosis. The accuracy and ease of linkage based diagnosis will be improved by the identification of the PKD1 gene as a microsatellite lies in the 3' untranslated region of this gene (KG-8) and several CA repeats are located 5' of the gene (see Figure 1a and 6; Peral, et al., 1994; Snarey, et al., 1994).

Experimental Procedures

Clinical Details of Patients

Family 77

77-2 and 77-3 are 48 and 17 years old, respectively and have typical ADPKD. Both have bilateral polycystic kidneys and 77-2 has impaired renal function. Neither patient manifests any signs of TSC (apart from cystic kidneys) on clinical and ophthalmological examination or by CT scan of the brain.

77-4 is 13 years old, severely mentally retarded and has multiple signs of tSC including adenoma sebaceum, depigmented macules and periventricular calcification on CT scan. Renal ultrasound reveals a small number of bilateral renal cysts.

ADPKD patients

OX875 developed ESRD from ADPKD, aged 46. Progressive decline in renal function had been observed over 17 years; ultrasound examinations documented enlarging polycystic kidneys with less extensive hepatic cystic disease. Both kidneys were removed after renal transplantation and pathological examination showed typical advanced cystic disease in kidneys weighing 1920g and 340g (normal average 120g).

OX114 developed ESRD from ADPKD aged 54: diagnosis was made by radiological investigation during an episode of abdominal pain aged 25. A progressive decline in renal function and the development of hypertension was subsequently observed. Ultrasonic examination demonstrated enlarged kidneys with typical cystic disease, with less severe hepatic involvement.

OX32 is a member of a large kindred affected by typical ADPKD in which several members have developed ESRD. The patient himself has been observed for 12 years with progressive renal failure and hypertension following ultrasonic demonstration of polycystic kidneys.

No signs of TSC were observed on clinical examination of any of the ADPKD patients.

DNA Electrophoresis and Hybridisation

DNA extraction, restriction digests, electrophoresis, Southern blotting, hybridisation and washing were performed by standard methods or as previously described (Harris, et al., 1990). FIGE was performed with the Biorad FIGE Mapper using programme 5 to separate fragments from 25-50 kb. High

molecular weight DNA for PFGE was isolated in agarose blocks and separated on the Biorad CHEF DRII apparatus using appropriate conditions.

Genomic DNA probes and somatic cell hybrids

Many of the DNA probes used in this study have been described previously: MS205.2 (D16S309; Royle, et al., 1992); GGG1 (D16S259; Germino, et al., 1990); N54 (D16S139; Himmelbauer, et al., 1991); SM6 (D16S665), CW23, CW21, and JH1 (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). Microsatellite probes for haplotype analysis were KG8 and W5.2 (Snarey, et al., 1994)SM6, CW3 and CW2, (Peral, et al., 1994), 16AC2.5 (Thompson, et al., 1992); SM7 (Harris, et al., 1991), VK5AC (Aksentijevich, et al., 1993).

New probes isolated during this study were: JH4, JH5, JH6, 11 kb, 6 kb and 6 kb BamH I fragments, respectively, and JH13 and JH14, 4 kb and 2.8 kb BamH I-EcoR I fragments, respectively, all from the cosmid JH2A; JH8 and JH10 are 4.5 kb and 2 kb Sac I fragments, respectively and JH12 a 0.6 Sac I-BamH I fragment, all from JH4; 8S1 and 8S3 are 2.4 kb and 0.6 kb Sac II fragments, respectively, from JH8; CW10 is a 0.5 kb Not I-Mlu I fragment of SM25A; JH17 is a 2 kb EcoR I fragment of NM17.

The somatic cell hybrids N-OH1 (Germino, et al., 1990), P-MWH2A (European Chromosome 16 Tuberous Sclerosis Consortium, 1993) and Hy145.19 (Himmelbauer, et al., 1991) have previously been described. Somatic cell hybrids containing the paternally derived (BP2-10) and maternally derived (BP2-9) chromosomes from OX114 were produced by the

method of Deisseroth and Hendrick (1979).

Constructing a cosmid contig

Cosmids were isolated from chromosome 16 specific and total genomic libraries, and a contig was constructed using the methods and libraries previously described (European Chromosome 16 Tuberous Sclerosis Consortium, 1993). To ensure that cosmids were derived from the 16p13.3 region (not the duplicate 16p13.1 area) initially, probes from the single copy area were used to screen libraries (e.g. CW21 and N54). Two cosmids mapped entirely within the area duplicated, CW10III and JC10.2B. To establish that these were from the PKD1 area, they were restriction mapped and hybridised with the probe CW10. The fragment sizes detected were compared to results obtained with hybrids containing only the 16p13.3. are (Hy145.19) or only the 16p13.1 region (P-MWH2A).

FISH

FISH was performed essentially as previously described (Buckle and Rack, 1993). The hybridisation mixture contained 100 ng of biotin-II-dUTP labelled cosmid DNA and 2.5 mg human Cot-1 DNA (BRL), which was denatured and annealled at 37°C for 15 min prior to hybridisation at 42°C overnight. After stringent washes the site of hybridisation was detected with successive layers of fluorescein-conjugated avidin (5 mg/ml) and biotinylated ani-avidin (5 mg/ML) Vector Laboratories). Slides were mounted in Vectashield (Vector Laboratories) containing 1 mg/ml propidium iodide and 1 mg/ml 4', 6-diamidino-2-phenylindole

(DAPI), to allow concurrent G-banded analysis under UV light. Results were analysed and images captured using a Bio-Rad MRC 600 confocal laser scanning microscope.

cDNA screening and characterisation

Foetal brain cDNAs libraries in λ phage (Clonetech and Stratagene) were screened by standard methods with genomic fragments in the single copy area (equivalent to CW23 and CW21) or with a 0.8 kb Pvu II-Eco RI single copy fragment of AH3. Six PBP cDNAs were characterised: AH4 (1.7 kb) and 3A3 (2.0 kb) are described in European Chromosome 16 Tuberous Sclerosis Consortium, 1993, and four novel cDNAs AH3 (2.2 kb), AH6 (2.0 kb), A1C (2.2 kb) and B1E (2.9 kb). A Striatum library (Stratagene) was screened with JH4 and a HG-C cDNA, 11BHS21 (3.8 KB) WAS ISOLATED, 21p.9 is a 0.9 kb Pvu II-EcoR I subclone of this cDNA. A HG-A or HG-B cDNA, HG-4 (7 kb) was also isolated by screening the foetal brain library (Stratagene) with JH8. HG-4/1.1 is a 1.1 kb Pvu II-EcoR I fragment from the 3' end of HG-4. 1A1H.6 is a 0.6 kb Hind III-EcoR I subclone of a TSC2 cDNA, 1A-1 (1.7 kb), which was isolated from the Clonetech library. Each cDNA was subcloned into Bluescript and sequenced utilising a combination of sequential truncation and liigonucleotide primers using DyeDeoxy Terminators (Applied Biosystems) and an ABI 373A DNA Sequencer (Applied Biosystems) or by hand with 'Sequenase' T7 DNA polymerase OUSB).

RNA Procedures

Total RNA was isolated from cell lines and tissues by the method of Chomczynskiand Sacchi (1987) and enrichment for mRNA made using the PolyAT tract mRNA Isolation System (Promega). For RNA electrophoresis 0.5% agarose denaturing formaldehyde gels were used which were Northern blotted.

hybridised and washed by standard procedures. The 0.24 - 9.5 kb RNA (Gibco BRL) size standard was used and hybridisation of the probe (1-9B3) to the 13 kb Utrophin transcript (Love, et al., 1989) in total fibroblast RNA was used as a size marker for the large transcripts.

RT-PCR was performed with 2.5 mg of total RNA by the method of Brown et al. (1990) with random hexamer primers, except that AMV-reverse transcriptase (Life Sciences) was employed. To characterise the deletion of the PBP transcript in OX114 we used the primers:

AH# F95' TTT GAC AAG CAC ATC TGG CTC TC 3'

AH3 B75' TAC ACC AGG AGG CTC CGC AG 3'

in a DMSO containing PCR buffer (Dode, et al., 1990) with 0.5 mM $MgCl_2$ and 36 cycles of: 94°C, 1 min; 61°C, 1 min; 72°C, 2 min plus a final extension of 10 min. The 3A3 C primers used to amplify the OX32 cDNA and DNA were:

3A3 C15' CGC CGC TTC ACT AGC TTC GAC 3'

3A3 C25' ACG CTC CAG AGG GAG TCC AC 3'

These were employed in a PCR buffer and cycle previously described (Harris, et al., 1991) with 1mM $MgCl_2$ and an annealing temperature of 61°C.

PCR products for sequencing were amplified with Pfu-1 (Stratagene) and ligated into the Srf-1 site in PCR-Script (Stratagene) in the presence of Srf-1.

RNAse protection

Tissues from normal and end-stage polycystic kidneys were immediately homogenised in guanidinium thiocyanate. RNA was purified on a cesium chloride gradient and 30 mg

total RNA was assayed by RNAse protection by the method of Melton, et al., (1984) using a genomic template generated with the 3A3, C primers.

Heteroduplex Analysis

Heteroduplex analysis was performed essentially as described by Keen et al. (1991). Samples were amplified from genomic DNA with the 3A3, C primers, heated at 95°C for 5 minutes and incubated at room temperature for at least 30 minutes before loading on a Hydrolink gel (AT Biochem). Hydrolink gels were run for 12-18 hours at 250V and fragments observed after staining with ethidium bromide.

Extraction and amplification of paraffin-embedded DNA

DNA from formalin fixed, paraffin wax embedded kidney tissue was prepared by the method of Wright and Manos (1990), except that after proteinase K digestion overnight at 55°C, the DNA was extracted with phenol plus chloroform before ethanol precipitation. Approximately 50 ng of DNA was used for PCR with 1.5 mM $MgCl_2$ and 40 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 40 s, plus a 10 min extension at 72°C.

The oligonucleotide primers designed to amplify across the genomic deletion of OX875 were:

AHF2 : 5' - GGG CAA GGG AGG ATG ACA AG - 3'

JH14B3 : 5' - GGG TTT ATC AGC AGC AAG CGG - 3'

which produced a product of about 220 bp in individuals with the OX875 deletion.

3' RACE analysis of WS-212

3' RACE was completed essentially as described

(European Polycystic Kidney Disease Consortium (1994)). Reverse transcription was performed with 5µg total RNA with 0.5µg of the hybrid dT_{15} adapter primer using conditions previously described (Fronman et al. (1988)). A specific 3' RACE product was amplified with the primer F5 and adapter primer in 0.5mM $MgCl_2$ with the program: 57°C, 60s; 72°C, 15 minutes and 30 cycles of 95°C, 40s; 57°C, 60s; 72°C, 60s plus 72°C, 10 minutes. The amplified product was cloned using the TA cloning system (Invitrogen) and sequenced by conventional methods.

Genomic and cDNA Probes and somatic cell hybrids

The genomic clones CW21, JH5, JH6, JH8, JH10. JH12, JH13 and JH14 and the cDNAs A1C, AH3, 3A3 and AH4 are described herein. Newly described probes are: SM3 a 2.0 kb BamH 1 subclone of the cosmid SM11, JH9, 2.4kb Sac 1 fragment and JH11, 1.2kn Sac 1 - BamH1 fragment, both from JH4. See Eur. Polycystic Kidney Disease Consortium, 1994 and Eur. Chromosome 16 Tuberous clerosis Consortium 1993 for all above clones. DFS5 is a 4.2 kb Not 1-Hind 111 fragment of CW23 (Eur. Chromosome 16 Tuberous Sclerosis Consortium, 1993). The cDNAs; BPG4, BPG6, BPG7C and 13-A were isolated from a fetal brain cDNA library in λ phage (Stratagene) and are 7 kb, 2 kb, 4.5 kb and 1.2 kb respectively.

The somatic cell hybrids have previously been described, P-MWH2A (Eur. Chromosome 16 Tuberous Sclerosis Consortium, 1993) and Hy145.19 (Himmelbauer et al., 1991).

Exon linking

Total cellular RNA from the radiation hybrid Hy145.19

was reverse transcribed using random hexamers (Eur. Polycystic Kidney Disease Consortium, 1994). This material was used as a template for PCR using the proof reading polymerase *Pfu*-1 with the primer pairs described in Table 2. The resultant products were cloned into the *Srf*-1 site of pPCRscript (SK+) plasmid.

Sequencing

Full length sequence was obtained from the genomic clones, HG cDNAs and exon link clones using the progressive unidirectional deletion technique of Henikoff, (1984). Both strands were then sequenced using DyeDeoxy Terminator Cycle Sequencing and an Applied Biosystems Sequencer 373A. Contig assembly was done using the programmes Assembly line (vs 1.0.7), SeqEd (vs 1.03) and MacVector (4.1.4).

Primer Extension

Primer extension was performed on total cellular fibroblast RNA. 25µg of RNA was annealed at 60°C in the presence of 400mM NaCl to 0.01pM of HPLC pure oligonucleotide which had been end labelled to a specific activity of 3 x 10^7 cpm/pM with ^{32}P. Primer extension was then performed in the presence of 50mM Tris pH8.2, 10mM DTT, 6mM $MgCl_2$, 25mg/ml Actinomycin D. 0.5mM dNTPs, and 8 units of AMV reverse transcriptase. The extension reaction was continued for 60 min at 42°C. The extension products were compared to a sequencing ladder generated using the same primer on the genomic clone SM3. The primers used were:

N2765:5'-GGCGCGGCGGGCGGCATCGTTAGGGCAGCG-3'

N5496:5'-GGCGGGCGGCATCGTTAGGGCAGCGCGCGC-3'

N5495:5'-ACCTGCTGCTGAGCGACGCCCGCTCGGGGC-3'.

Analysis of sequence homology

The predicted PKD1 protein was analyzed for homologies with known proteins in the SwissProt and NBRF database using the BLAST (Altschul et al., 1990) and FASTA (Pearson et al., 1988) algorithms. Layouts were prepared by hand and using the programme Pileup.

Transmembrane regions

Potential transmembrane segments were identified by the method of Sipos and von Heljne (Sipos et al., 1993), using the GES hydrophobicity scale (Engelmen et al., 1986) and a trapezoid sliding window (a full window of 21 residues and a core window of 11 residues), as recommended. Candidate transmembrane domains were selected on the basis of their average hydrophobicity <H>, and were classified as certain (<H> ≥ 1.0) or putative (0.6 , <H> <1).

The best topology for the protein was predicted on the basis of three different criteria: a) the net charge difference between the 15 N-terminal and the 15 C-terminal residues flanking the most N-terminal transmembrane segment (Hartmann et al., 1989); b) the difference in positively charged residues between the two sides of the membrane in loops smaller than 60 residues, and c) the analysis of the overall amino acid composition of loops longer than 60 residues by the compositional distance method (Nakashima et al., 1992). Using the above criteria the TopPred II program (Sipos wt al., 1993) calculated all the possible topologies of the proteins including the certain transmembrane segments

and either included or excluded each of the putative segments to determine the most likely structure.

PKD1 Protein Purification

The PKD1 protein may be purified according to conventional protein purification procedures well known in the art. Alternatively, the protein may be purified from cells harboring a plasmid containing an expressible PKD1 gene. For example, the protein may be expressed in an E.coli expression system and purified as follows.

Cells are grown in a 10 liter volume in a Chemap Fermentor (Chemapec, Woodbury, NY) in 2% medium. Fermentation temperature may be 37'C, pH 6.8, and air as provided at 1 vvm. Plasmid selection may be provided using ampicillin for a plasmid containing an ampicillin resistance gene. Typical yield (wet weight) is 30 g/l.

For cell lysis, 50g wet cell weight of E.coli containing the recombinant PKD1 plasmid may be resuspended in a final volume of 100ml in 50 mM Tris-HCl pH 8.0, 5 mM EDTA, 5mM DTT, 15 mM mercaptoethanol, 0.5% triton X-100, and 5 mM PMSF. 300 mg lysozyme is added to the suspension, and incubated for 30 min at room temperature. The material is then lyzed using a BEAD BEATER (R) (Biospec Products, Bartlesville, OK) containing an equal volume of 0.1-0.15 um glass beads. The liquid is separated from the beads and the supernatant removed, the pellet dissolved in 20 mM Tris-Cl pH 8.0.

The protein may be purified from the supernatant using DEAE chromatography, as is well known in the art.

Preparation of Antibodies.

Antibodies specific for PDK1 protein or a fragment thereof are prepared as follows. A peptide corresponding to at least 8 amino acid residues of the PKD1 sequence of Fig. 15, are synthesized. Coupling of the peptide to carrier protein and immunizations is performed as described (Dymecki, S.M., J. Biol. Chem 267:4815-4823, 1992). Rabbit antibodies against this peptide are raised and sera are titered against peptide antigen by ELISA. The sera exhibiting the highest titer (1:27,000) are most useful.

Techniques for preparing monoclonal antibodies are well known, and monoclonal antibodies of this invention may be prepared by using the synthetic polypeptides of this invention, preferably bound to a carrier, as the immunogen as was done by Arnheiter et al., Nature, 294, 278-280 (1981).

Monoclonal antibodies are typically obtained from hybridoma tissue cultures or from ascites fluid obtained from animals into which the hybridoma tissue was introduced. Nevertheless, monoclonal antibodies may be described as being "raised to" or "induced by" the synthetic polypeptides of this invention or their conjugates with a carrier.

Antibodies are utilized along with an "indicating group" also sometimes referred to as a "label". The indicating group or label is utilized in conjunction with the antibody as a means for determining whether an immune reaction has taken place, and in some instances for determining the extent of such a reaction.

The indicating group may be a single atom as in the case of radioactive elements such as iodine 125 or 131, hydrogen 3 or sulfur 35, or NMR-active elements such as fluorine 19 or nitrogen 15. The indicating group may also be a molecule such as a fluorescent dye like fluorescein, or an enzyme, such as horseradish peroxidase (HRP), or the like.

The terms "indicating group" or "label" are used herein to include single atoms and molecules that are linked to the antibody or used separately, and whether those atoms or molecules are used alone or in conjunction with additional reagents. Such indicating groups or labels are themselves well-known in immunochemistry and constitute a part of this invention only insofar as they are utilized with otherwise novel antibodies, methods and/or systems.

Detection of PKD1 and Subcellular Localization.

Another embodiment of this invention relates to an assay for the presence of PKD1 protein in cells. Here, an above-described antibody is raised and harvested. The antibody or idiotype-containing polyamide portion thereof is then admixed with candidate tissue and an indicating group. The presence of the naturally occurring amino acid sequence is ascertained by the formation of an immune reaction as signaled by the indicating group. Candidate tissues include any tissue or cell line or bodily fluid to be tested for the presence of PKD1.

Metabolic labeling, immunoprecipitation, and immunolocalization assays are performed in cells as

described previously (Furth, M.E., et al., Oncogene 1:47-58, 1987; Laemmli, U.K., Nature 227:680-685, 1970; Yarden, Y., et al., EMBO J. 6:3341-3351, 1987; Konopka, J.B., et al., Mol. Cell. Biol. 5:3116-3123, 1985). For immunoblot analysis, total lysates are prepared (using Fruth's lysis buffer) (Fruth, M.E., et al., Oncogene, 1:47-58, 1987). Relative protein concentrations are determined with a colorimetric assay kit (Bio-Rad) with bovine serum albumin as the standard. A protein of lysate containing approximately 0.05 mg of protein is mixed with an equal volume of 2 x SDS sample buffer containing 2 mercaptoethanol, boiled for 5 min., fractioned on 10% polyacrylamide-SDS gels (Konopka, J.B., et al., J.Virol., 51:223-232, 1984) and transferred to immunobilon polyvinyldine difluoride (Millipore Corp., Bedford, MA) filters. Protein blots are treated with specific antipeptide antibodies (see below). Primary binding of the PKD1-specific antibodies is detected using anti-IgG second antibodies conjugated to horseradish peroxidase and subsequent chemiluminescence development ECL Western blotting system (Amersham International).

For metabolic labeling, 10^6 cells are labeled with 100 µCi of ^{35}S-methionine in 1 ml of Dulbecco's modified Eagles medium minus methionine (Amersham Corp.) for 16h. Immunoprecipitation of PKD1 protein from labeled cells with antipeptide antiserum is performed as described (Dymecki, S.M., et al., supra). Portions of lysates containing 10^7 cpm of acid-insoluble ^{35}S-methionine are incubated with 1 µg of

the antiserum in 0.5 ml of reaction mixture. Immunoprecipitation samples are analyzed by SDS-polylarcylamide gel electrophoresis and autoradiography.

For immunolocalization studies, 10^7 CMK cells are resuspended in 1 ml of sonication buffer (60mM Tris-HCl, pH 7.5, 6 mM EDTA, 15 mM EGTA, 0.75M sucrose, 0.03% leupeptin 12mM phenylmethylsulfonyl fluoride, 30 mM 2-mercaptoethanol). Cells are sonicated 6 times for 10 seconds each and centrifuged at 25,000 xg for 10 min at 4°C. The pellet is dissolved in 1 ml of sonication buffer and centrifuged at 25,000 x g for 10 min at 4°C.

The pellet (nucleus fraction) is resuspended in 1 ml of sonication buffer and added to an equal volume of 2 x SDS sample buffer. The supernatant obtained above (after the first sonication) is again centrifuged at 100,000 x g for 40 min at 4°C. The supernatant (cytosolic fraction) is removed and added to an equal volume of 2 x concentrated SDS sample buffer. The remaining pellet (membrane fraction) is washed and dissolved in sonication buffer and SDS sample buffer as described above. Protein samples are analyzed by electrophoresis on 10% polyacrylamide gels, according to the Laemmli method (Konopka, J.B., supra). The proteins are transferred from the gels on a 0.45-µm polyvinylidine difluoride membrane for subsequent immunoblot analysis. Primary binding of the PKD1 specific antibodies is detected using anti-IgG second antibodies conjugated to horseradish peroxidase.

For immunohistochemical localization of PKD1 protein,

CMK cells or U3T3 are grown on cover slips to approximately 50% confluence and are washed with PBS (pH 7.4) after removing the medium. The cells are prefixed for 1 min at 37°C in 1% paraformaldehyde containing 0.075% Triton X-100, rinsed with PBS and then fixed for 10 min with 4% paraformaldehyde. After the fixation step, cells are rinsed in PBS, quenched in PBS with 0.1 and finally rinsed again in PBS. For antibody staining, the cells are first blocked with a blocking solution (3% bovine serum albumin in PBS) and incubated for 1 h at 37°C. The cells are then incubated for 1 h at 37°C with antiserum (1:100 dilution or with preimmune rabbit serum (1:100). After the incubation with the primary antibody, the cells are washed in PBS containing 3% bovine and serum albumin and 0.1% Tween 20 and incubated for 1 h at 37 C in fluorescein-conjugated donkey anti-rabbit IgGs (Jackson Immunoresearch, Maine) diluted 1:100 in blocking solution.

The coverslips are washed in PBS (pH 8.0), and glycerol is added to each coverslip before mounting on glass slides and sealing with clear nail polish. All glass slides are examined with a Zeiss Axiophot microscope.

An indicating group or label is preferably supplied along with the antibody and may be packaged therewith or packaged separately. Additional reagents such as hydrogen peroxide and diaminobenzideine may also be included in the system when an indicating group such as HRP is utilized. Such materials are readily available in commerce, as are many indicating groups, and need not be supplied along with

the diagnostic system. In addition, some reagents such as hydrogen peroxide decompose on standing, or are otherwise short-lived like some radioactive elements, and are better supplied by the end-user.

Pharmaceutical Compositions of the Invention; Dosage and Administration

Pharmaceutical formulations comprising PKD1 nucleic acid or protein, or mutants thereof, can be prepared by procedures well known in the art. For example, as injectables, e.g., liquid solutions or suspensions. Solid forms for solution in, or suspension in, a liquid prior to injection also can be prepared. Optionally, the preparation also can be emulsified. The active ingredient can be mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. For example, water, saline, dextrose, glycerol, ethanol, etc. or combinations thereof. Also useful are wetting or emulsifying agents, pH buffering agents or adjuvants. PKD1 protein or DNA can be administered parenterally, by injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories and, in some cases, oral formulations. In each case, the active protein or the nucleic acid will be present in the range of about 0.05% to about 10%, preferably in ther ange of about 1-2% by weight. Alternatively, the active protein or the nucleic acid will be administered at a dosage of about 10mg-2kg/kg body weight, preferably 50mg-400mg/kg/body weight. Administration may be daily, weekly,

or in a single dosage, as determined by the physician.

OTHER EMBODIMENTS

Other embodiments will be evident to those of skill in the art. It should be understood that the foregoing detailed description is provided for clarity only and is merely exemplary. The spirit and scope of the present invention are not limited thereto, being defined by the claims set forth below.

REFERENCES

Aksentijevich et al., Am. J. Hum. Genet. 53:451-461, (1993).

Altschul et al., J. Mol. Biol. 215:403-410, (1990).

Bevilacqua, M.P., et al., Science 243:1160-1165, (1989).

Bork et al., Protein Science 2:1185-1187, (1993).

Breuning et al., Lancet ii, 1359-1361, (1987).

Breuning et al., J. Med. Genet. 27:603-613, (1990).

Brook-Carter et al., Nature Genetics 8:328-332, (1994).

Brown et al., Nucl. Acids Res. 18:4191-4195, (1990).

Brümmendorf, T., et al., Protein Profile 1:951-1058, (1994).

Buckle et al., Human Genetic Disease Analysis; IRL Press (K.E. Davies, Ed.) 2:59-82, Oxford, (1993).

Carone, F.A., et al., Laboratory Investigations 70:437-448, (1994).

Carone, F.A., et al., Kidney International 47:861-868, (1995).

Calvet. J.P., Kidney International 43:101-108, (1993).

Chapman et al., N. Eng. J. Med. 327:916-920, (1992).

Chao. M.V., Neuron 9:583-593, (1992).

Chomczynski et al., Anal. Biochem. 162:156-159, (1987).

Curtis et al., Proc. of the Nat'l. Acad. of Sci., USA 89:8356-8360, (1992).

Dalgaard, O.Z., Acta Medica Scandinavica 158:1-251, (1957).

Daoust, M.C., et al., Genomics 25:733-736, (1995).

Davies et al., Q.J. Med. 79:477-485, (1991).

Deisseroth et al., Proc. Natl. Acad. Sci. USA 76:2185-2189, (1979).

Dode et al., Brit. J. Haemat. 76:275-281, (1990).

Drickamer, K., Kidney Int'l. 32:167-180, (1987).

Drickamer, K., J. Biol. Chem. 263:9557-9560, (1988).

Ekblom, P., FASEB Journal 3:2141-2150, (1989).

Engelman et al., Ann. Rev. Bioph. Chem. 15:321-353, (1986).

European Polycystic Kidney Disease Consortium, Cell 77:881-894, (1994).

European Chromosome 16 Tuberous Sclerosis Consortium, Cell 75:1305- 1315, (1993).

Fink et al., J. Amer. Soc. Nephrology 3:1863-1870, (1993).

Fink et al., Kidney Int. 45:1153-1162, (1994).

Fronman et al., Biochemistry 85:8998-9002, (1988).

Gabow, P.A., Kidney Int. 40:989-996, (1991).

Gabow, P.A., N.E. J. of Medicine 329:332-342, (1993).

Gabow, P.A., Amer. J. of Kidney Diseases 16:403-413, (1990).

Germino et al., Am. J. Hum. Genet. 46:925-933, (1990).

Germino et al., Genomics 13:144-151, (1992).

Gower, H.J., et al., Cell 55:955-964, (1988).

Green et al., Nature Genet. 6:193-196, (1994).

Harpaz, Y., et al., J. of the Mol. Biol. 238:528-539 (1994).

Harris et al., Genomics 7:195-206, (1990).

Harris et al., Lancet 338:1484-1487, (1991).

Hartmann et al., Proc. Nat'l. Acad. Sci. USA 86:5786-5790, (1989).

Henikoff, S., Gene 28:351-359, (1984).

Himmelbauer et al., Amer. J. Human Genetics 48:325-334, (1991).

Hossack et al., N. Eng. J. Med. 319:907-912, (1988).

Huston et al., J. Amer. Soc. of Nephrology 3:1871-1877,

(1993).

Hyland et al., Hum. Genet. 84:286-288, (1990).

Jia, R., et al., J. of Biol. Chem. 269:1839-1844 (1994).

Jones, E.Y., et al., Nature 373:539-544, (1995).

Keen et al., Trend Genet. 7:5, (1991).

Kimberly, W.J., et al., Genomics 18:467-472, (1993).

Kimberling et al., N. Eng. J. Med. 319:913-918, (1988).

Kobe et al., Trends in Bioch. Sci. 19:415-421, (1994).

Kornblihtt, A.R., et al., EMBO Journal 4:1755-1759, (1985).

Kozak, M., Nucleic Acids Res. 15:8125-8148, (1987).

Kuma et al., Mol. Biol. and Evolution 10:539-551, (1993).

Kwon, B.S., et al., Proc. of the Nat'l. Acad. of Sci., USA 88:9228- 9232, (1991).

Lamballe et al., Cell 66:967-979, (1991).

Legius et al., Nature Genet. 3:122-126, (1993).

Love et al., Nature 339:55-58, (1989).

Mandel, J-L, Nature Genetics 4:8-9.

Matsushita, O., et al., Journal of Bacteriology 176:149-156, (1994).

McFarland, K.C., et al., Science 245:494-499, (1989).

Melton et al., Nuc. Acid Res. 12:7035-7056.

Milutinovic, J., et al., Amer. J. of Med. 68:741-744, (1980).

Milutinovic, J., et al., Amer. J. of Clin. Path. 73: 740-747, (1979).

Nakashima et al., FEBS Letters 303:141-146, (1992).

Oldberg, et al., EMBO J. 8:2601-2604, (1989).

Oldberg et al., Biochemical J. 243:255-259, (1987).

Parfrey et al., N. Eng. J. Med. 323:1085-1090, (1990).

Pearson et al., Proc. Nat'l Acad. Sci. USA 85:2444-2448, (1988).

Peral et al., Am. J. Hum. Genet. 54:899-908.

Peral et al., Human Molecular Genetics (in press), (1995).

Peters, D.J.M., et al., Nature Genetics 5:359-362, (1993).

Peters, D.J.M., et al., Contributions to Nephrology: Polychystic Kidney Disease (eds. Breuning, M.H., Devoto, M. & Romeo, G), p. 128-139 (1992).

Pound et al., J. Med. Genet. 29:247-248, (1992).

Ravine et al., Lancet 337:127-129, (1991).

Ravine D., et al., Lancet 340:1330-1333, (1992).

Reeders, S.T., Nature Genet. 1:235-237, (1992).

Reeders et al., Lancet i, 6-8, (1986).

Reeders et al., Nature 317:542-544, (1985).

Reeders et al., Genomics 3:150-155. (1988).

Romeo et al., Lancet ii, 8-10, (1988).

Roth, G.J., Blood 77:5-19, (1991).

Rothberg et al., Genes and Development 4:2169-2187, (1990).

Royle et al., Nucl. Acids Res. 20:1164, (1992).

Ryynanen et al., J. Med. Genet. 24:462-465, (1987).

Schäfer, K., et al., Kidney International 46:134-152, (1994).

Scheff et al., Ann. Intern. Med. 92:202-204, (1980).

Sipos et al., European J. Biochemistry 213:1333-1340, (1993).

Snarey et al., Am. J. Hum. Genet. (in press), (1994).

Somlo et al., Genomics 13:152-158, (1992).

Somlo, S., et al., J. of the Amer. Soc. of Nephrology 4: 1371-1378, (1993).

Streuli, M., et al., Journal of Experimental Medicine 168:1523-1530, (1988).

Takagi et al., J. Bioch. Chem. 265:19721-19727, (1990).

Taylor, M.E., et al., J. of Biol. Chem. 265:12156-12162, (1990).

Thompson et al., Genomics 13:402-408, (1992).

Volkmer H., et al., Journal of Cell Biology 118:149-161, (1992).

von Heijne, G., Nuc. Acids Res. 14:4683-4691, (1986).

Wieringa, B., et al., Cell 37:915-925, (1984).

Weis et al., Nature 360:127-134, (1992).

Williams, A.F., et al., Annual Review of Immunology 6:381-405 (1988).

Wilson, P.D., et al., Kidney International 39:450-463, (1991).

Wright et al., PCR Protocols: A Guide to Methods and Applications, 153-166, (1990).

Zerres et al., J. Med. Genet. 30:583-588, (1993).

CLAIMS

1. An isolated nucleic acid sequence comprising:-

(a) a PKD1 gene or its complementary strand,

(b) a sequence substantially homologous to a substantial portion of a molecule defined in (a) above, or

(c) a fragment of a molecule defined in (a) or (b) above.

2. A sequence according to claim 1, wherein the PKD1 gene has the nucleic acid sequence according to Figure 15.

3. A sequence according to claim 1, wherein the PKD1 gene has the partial nucleic acid sequence according to Figure 7.

4. A sequence according to claim 1, wherein the PKD1 gene has the partial nucleic acid sequence according to Figure 10.

5. An isolated nucleic acid selected from the group consisting of:

(a) [OX114] a nucleic acid including a deletion of 446 base pairs between residues 1746-2192 as defined in Figure 7;

(b) [OX32] a nucleic acid including a deletion of 135 base pairs between residues 3696-3831 as defined in Figure 7;

(c) [OX875] a nucleic acid wherein about 5.5kb flanked by the two XbaI sites shown in Figure 3a are deleted and the

EcoR1 site separating the CW10 (41kb) and JH1 (18kb) fragments is thereby absent;

(d) (WS-53) a nucleic acid including a deletion of about 100kb encompassing the PKD1 gene, wherein the 3' end of the deletion lies between the JH1 and CW21 fragments and the 5' end of the deletion lies between the SM6 and JH17 frgments shown in Figure 6;

(e) (461) a nucleic acid wherein about 18 base pairs are deleted in the 75 base pair intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(f) (OX1054) a nucleic acid wherein about 20 base pairs are deleted in the 75 base pair intron amplified by the primer pair 3A3C insert at position 3696 of the 3' sequence as shown in Figure 11;

(g) (WS-212) a nucleic acid including a deletion of about 75kb downstream of the PKD1 gene and located between fragments SM9 and CW9 distal of the PKD1 gene and the PKD1 3'UTR proximal to the PKD1 gene as shown in Figure 12;

(h) (WS-215) a nucleic acid including a deletion of about 160kb encompassing the PKD1 gene, wherein the deletion extends 3' of the PKD1 gene to within fragment CW15 and 5' of the PKD1 gene to between fragments CW10 and CW36 as shown in Figure 12;

(i) (WS-227) a nucleic acid including a deletion of about 50kb encompassing the PKD1 gene, wherein the deletion extends 3' of the PKD1 gene to within fragment CW20 and 5' of the PKD1 gene to within fragment JH11 as shown in Figure

12;

(j) (WS-219) a nucleic acid including a deletion of about 27kb encompassing a portion of the PKD1 gene, wherein the deletion extends 3' of the PKD1 gene within fragment JH1 and into the PKD1 gene to within fragment JH6 as shown in Figure 12;

(k) (WS-250) a nucleic acid including a deletion of about 160kb encompassing the PKD1 gene, wherein the deletion extends 3' of the PKD1 gene to within fragment CW20 and 5' of the PKD1 gene to within fragment BLu24 as shown in Figures 1a and 12; and

(l) (WS-194) a nucleic acid including a deletion of about 65kb encompassing the PKD1 gene, wherein the deltion extends 3' of the PKD1 gene to within fragment CW20 and 5' of the PKD1 gene to within fragment CW10.

6. An isolated nucleic acid according to any preceding Claim, wherein the molecule is an RNA transcript comprising a sequence complementary to the coding region of the nucleic acid sequence according to Fig. 15 and comprising a length of about 14 KB.

7. An isolated nucleic acid according to claim 5 comprising an RNA transcript.

8. An isolated nucleic acid according to claim 6 comprising an RNA transcript.

9. A nucleic acid probe comprising 10 nucleotides complementary to 10 consecutive nucleotides of the PKD1 sequence according to Figure 15.

10. A nucleic acid probe according to claim 9 wherein said probe is between 15 nucleotides and 14 kb in length.

11. A nucleic acid probe according to claim 10, said probe being between 100 nucleotides and 5 kb in length.

12. A recombinant expression vector comprising the isolated nucleic acid according to claim 1.

13. A host cell comprising the vector of claim 12.

14. A recombinant expression vector comprising the isolated nucleic acid according to claim 5.

15. A recombinant expression vector comprising the isolated nucleic acid according to claim 7.

16. An isolated polypeptide comprising a PKD1 protein having the amino acid sequence according to Fig. 15.

17. An isolated polypeptide comprising a PKD1 protein fragment having the amino acid sequence according to Fig. 7.

18. An isolated polypeptide comprising a PKD1 protein

fragment having the amino acid sequence according to Fig. 10.

19. An isolated polypeptide comprising a PKD1 protein fragment having an amino acid sequence comprising the amino acid sequence according to Fig. 7 and the amino acid residue deletions defined by the nucleotide deletions of claim 5, parts (a), (b) and (j).

20. An immunoglobulin molecule having specificity for PKD1 protein, said protein comprising the amino acid sequence according to any one of Figures 7, 10 or 15.

21. A method for screening a subject to determine whether said subject is a PKD1-associated disorder carrier or has a PKD1-associated disorder, which method comprises detecting the presence or absence of PKD1 nucleic acid in a biological sample from said subject, wherein detection of a mutant or absent PKD1 nucleic acid is indicative of a PKD1-associated disorder.

22. A method for screening a subject to determine whether said subject is a PKD1-associated disorder carrier or has a PKD1-associated disorder, which method comprises detecting the presence or absence of PKD1 polypeptide in a biological sample from said subject, wherein detection of a mutant or absent PKD1 polypeptide is indicative of a PKD1-associated disorder.

23. A method according to claim 21, comprising detecting a genomic fragment comprising the PKD1 gene or a portion thereof, a genomic fragment comprising a flanking region of the PKD1 gene or PKD1 RNA.

24. A method according to claim 23, wherein said detection comprises hybridizing a PKD1 nucleic acid probe to nucleic acid from said biological sample and comparing the results thereof with results obtained using a biological sample from a subject who is not a carrier of a PKD1-associated disorder.

25. A method according to claim 25, wherein said detection includes applying a nucleic acid amplification process to said nucleic acid to amplify a fragment of the PKD1 nucleic acid.

26. A method according to claim 26, wherein said nucleic acid amplification process comprises amplifying a fragment of PKD1 nucleic acid utilizing a set of primers selected from the group consisting of:-

AH3 F9 : AH3 B7

3A3 C1 : 3A3 C2

AH4 F2 : JH14 B3.

27. A method according to claim 24 wherein said detection step comprises digesting nucleic acid from said biological sample to EcoR1 fragments and hybridising with a DNA probe

which hybridises to the restriction fragment in Figure 3(a) or 12.

28. A method according to claim 27, wherein nucleic acid from said biological sample is digested with EcoR I and said DNA probe is selected from the group consisting of the probes CW10, JH14, JH5, JH6, JH4, JH13, JH8, JH11 and CW36 identified in Figures 3a and 12.

29. A method according to claim 28 which comprises digesting said nucleic acid to provide BamH I fragments and hybridising with a DNA probe which hybridises to the BamH I fragment identified (B) in Figure 3(a).

30. A method according to claim 30, wherein said DNA probe comprises the DNA probe 1A1H0.6 identified herein.

31. A method of treating a patient afflicted with a PKD1-associated disorder comprising administering a nucleic acid sequence according to any of claims 1 to 8.

32. A method of treating or preventing a PKD1-associated disorder which method comprises administering to a patient in need thereof a PKD1 gene having the sequence according to Figure 15 so as to permit expression of PKD1 protein.

33. A method of treating or preventing a PKD1-associated disorder which method comprises administering to a patient

in need thereof a mutated PKD1 gene isolated from WS212 DNA so as to permit expression of PKD1 protein.

34. A diagnostic kit for amplifying a portion of the PKD1 gene, comprising a pair of nucleic acid primers complementary to a portion of the PKD1 nucleic acid sequence according to Fig. 15, and packaging means therefore.

35. A diagnostic kit according to claim 34, wherein the nucleic acid primers comprise one or more of the following sets:

AH3 F9 : AH3 B7;

3A3 C1 : 3A3 C2; and

AH4 F2 : JH14 B3.

36. A diagnostic kit for carrying out a method for determining whether said subject is a PKD1-associated disorder carrier or a patient having a PKD1-associated disorder, which kit includes a nucleic acid probe capable of hybridising to a sequence according to claim 1.

37. A diagnostic kit for carrying out a method for determining whether said subject is a PKD1-associated disorder carrier or a patient having a PKD1-associated disorder, which kit includes a nucleic acid probe capable of hybridising to a sequence according to claim 6 and packaging means therefore.

38. A diagnostic kit for carrying out a method for determining whether said subject is a PKD1-associated disorder carrier or a patient having a PKD1-associated disorder, which kit includes a nucleic acid probe capable of hybridising to a sequence according to claim 5 and packaging means therefore.

39. A diagnostic kit for detecting PKD1 nucleic acid, including the DNA probe CW10 and packaging means therefore.

40. A diagnostic kit for detecting PKD1 nucleic acid, including the DNA probe 1A1H0.6 and packaging means therefore.

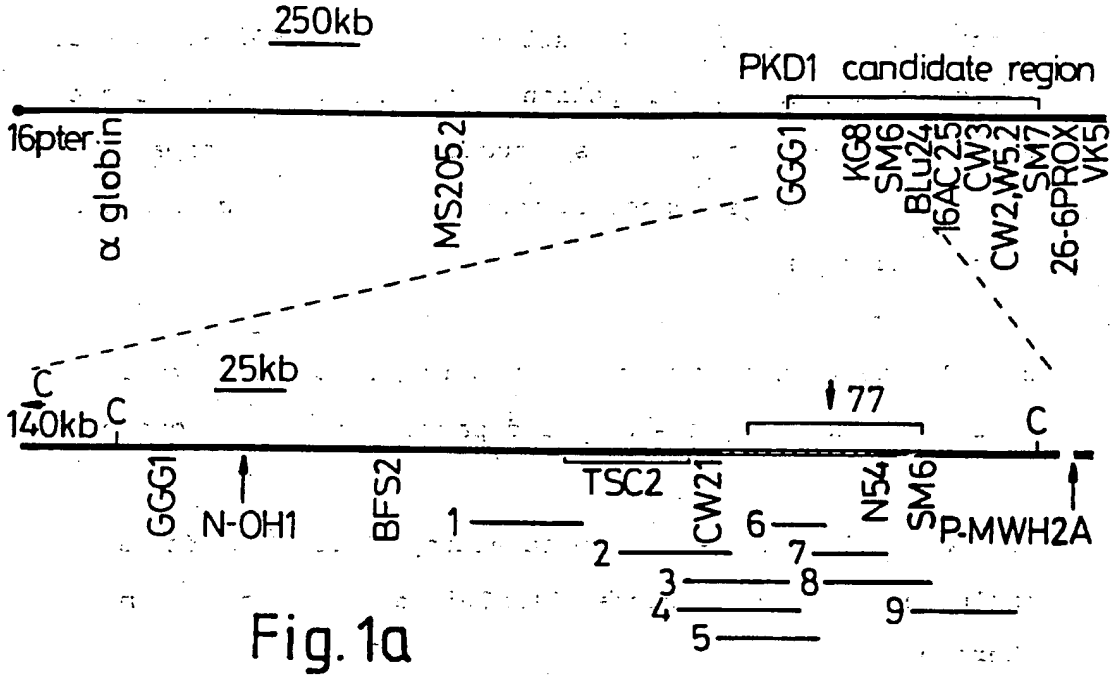

Fig. 1a

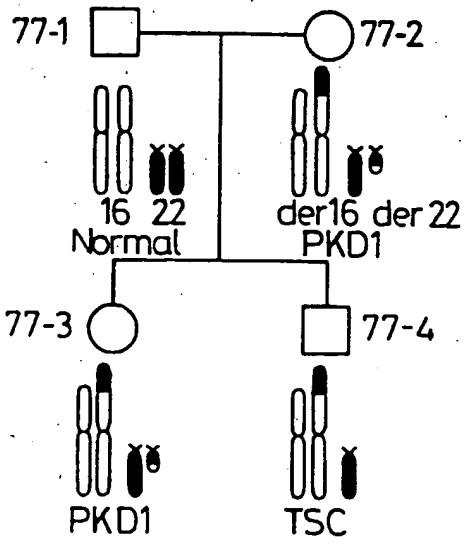

Fig. 1b

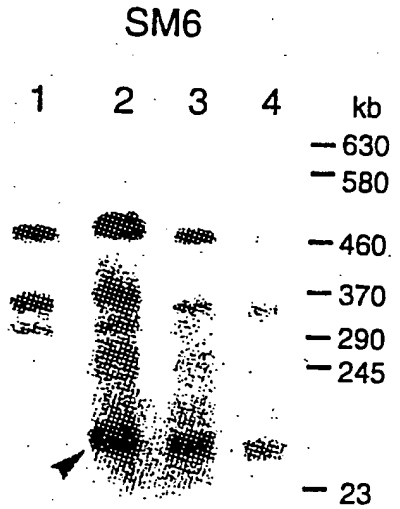

Fig. 1c

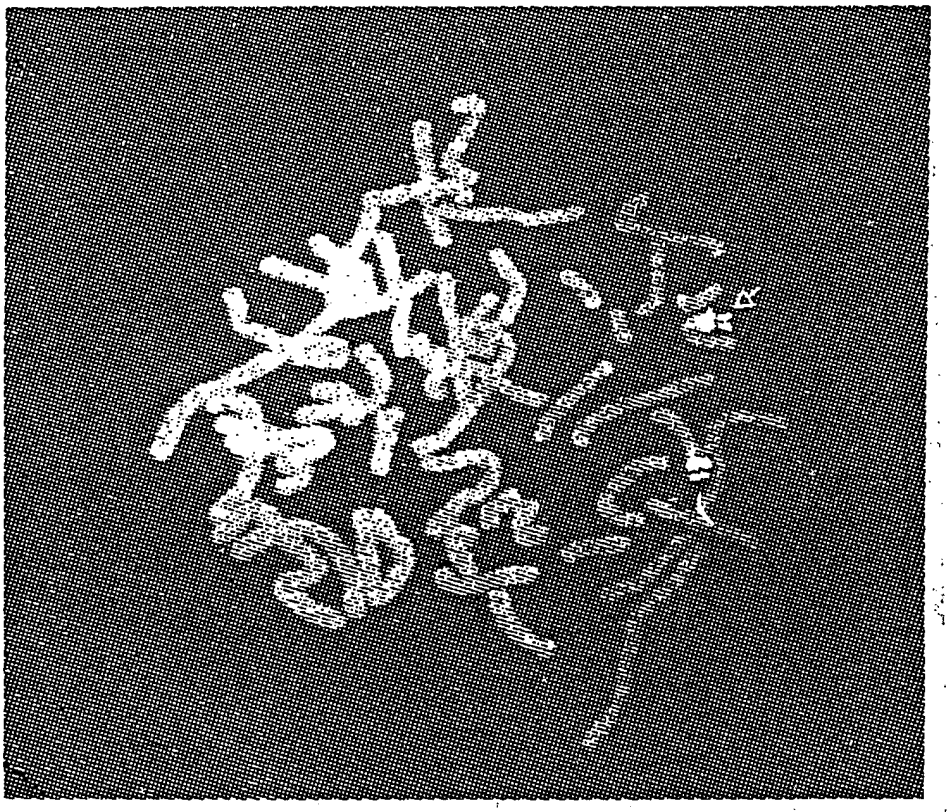

Fig. 2

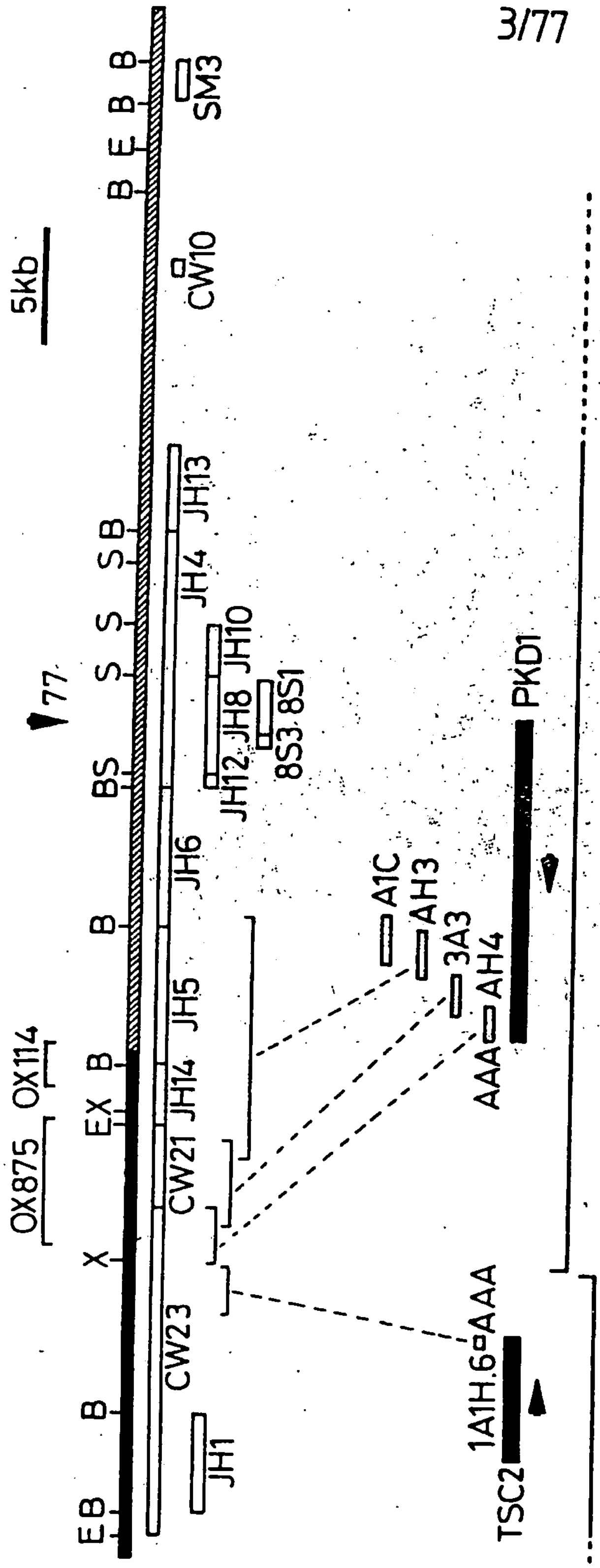
5kb
77
OX875
OX114
E B B X EX B B BS S S S S B B E B B
CW23
CW21
JH14
JH5
JH6
JH12
JH8
JH10
JH4
JH13
CW10
SM3
JH1
8S3
8S1
A1C
AH3
3A3
AH4
AAA
1A1H.6
AAA
TSC2
PKD1

Fig. 3a

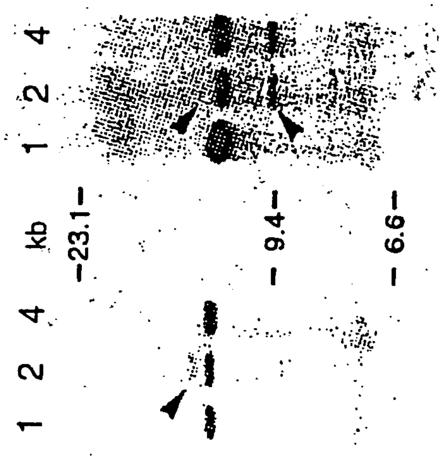
8S3
1 2 4
8S1
1 2 4
kb
-23.1-
- 9.4-
- 6.6-

Fig. 3b

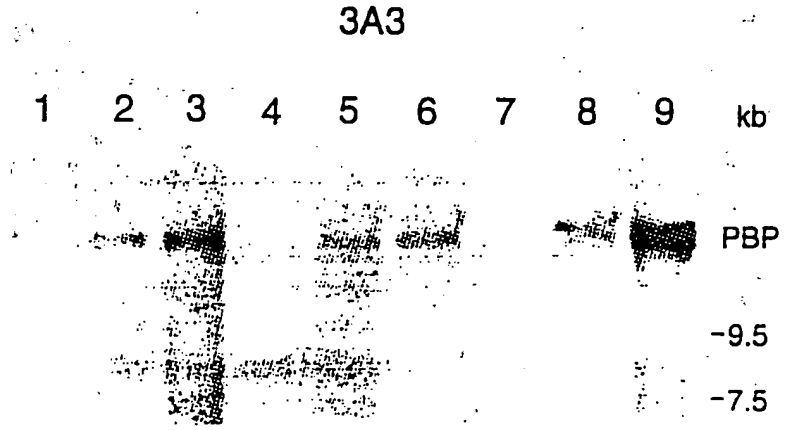

Fig.4a

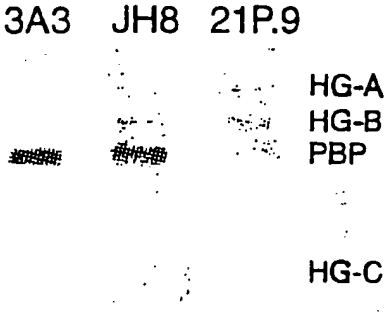

Fig.4b

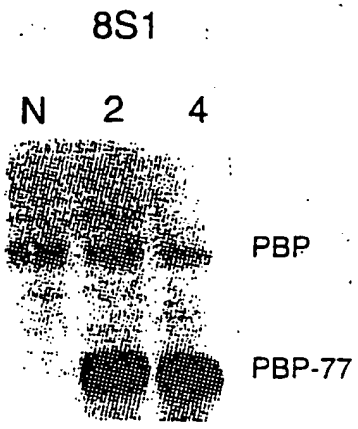

Fig.4c

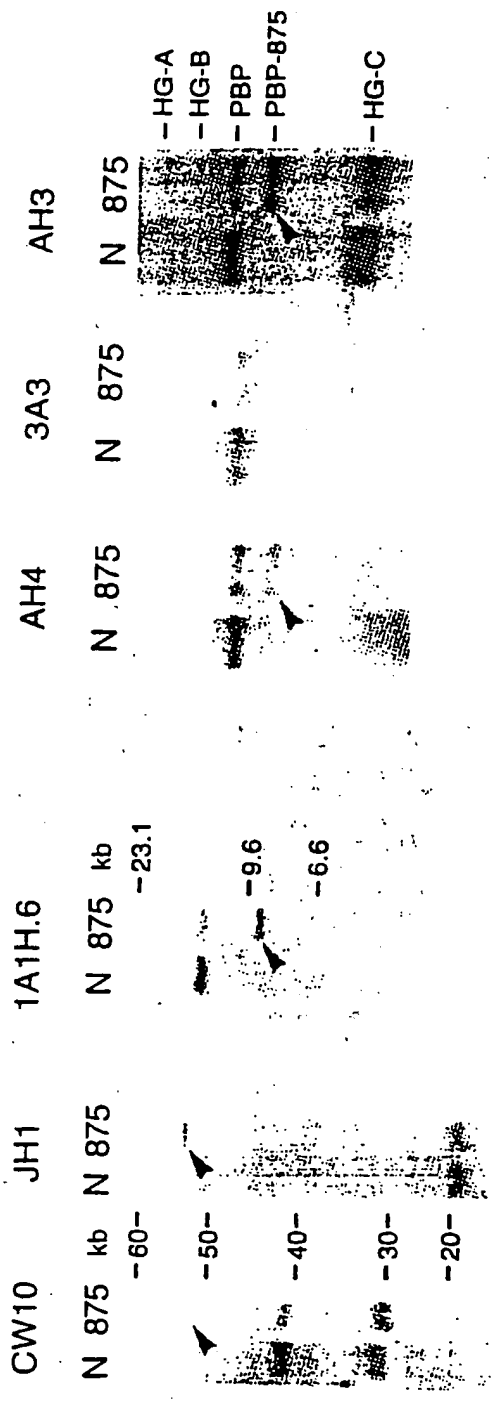

Fig. 5a

Fig. 5b

Fig. 5d

Fig. 5c

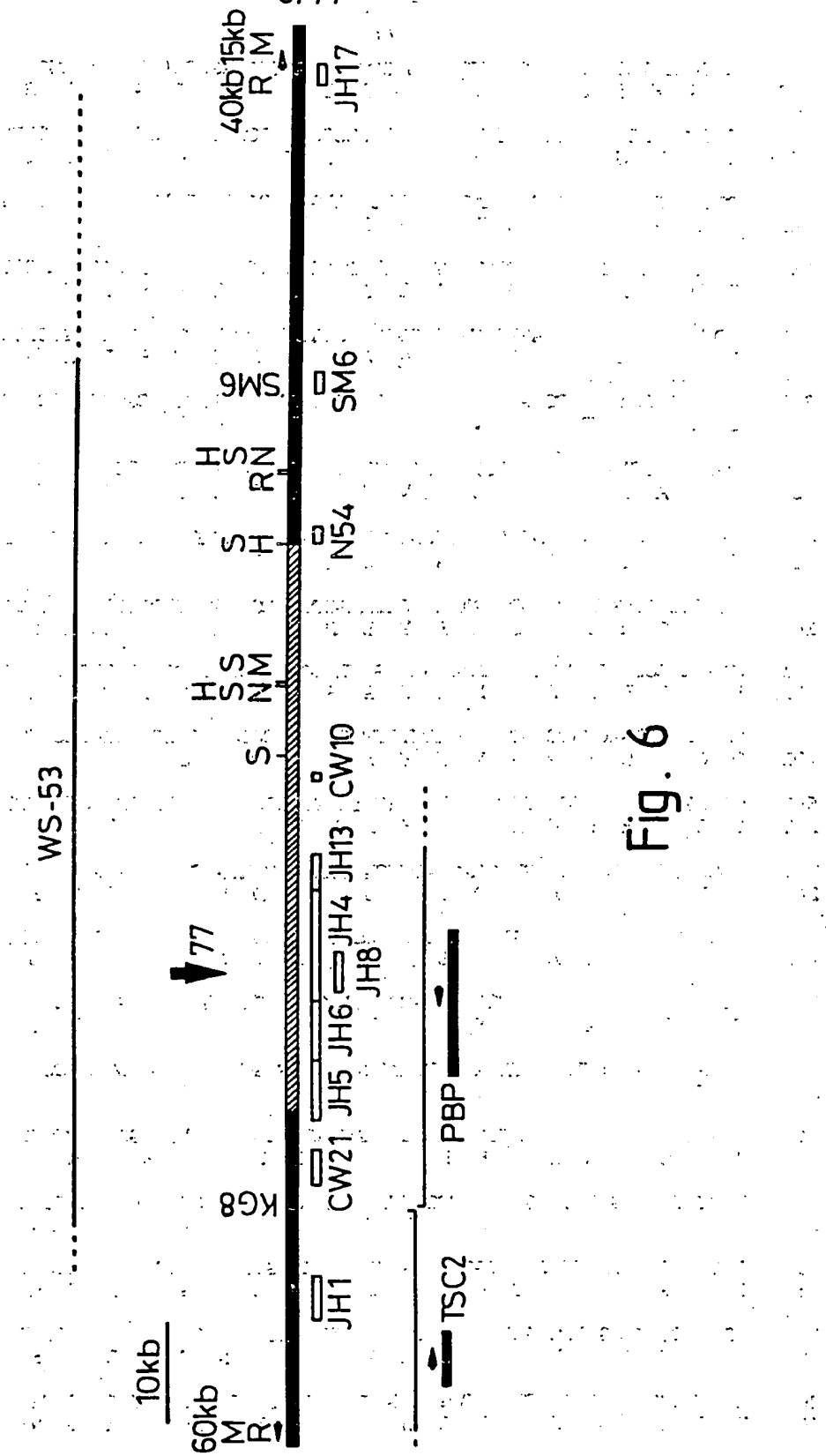
40kb 15kb
R M
JH17
9WS
SM6
H S
R N
S H
N54
H S S
N M
S
CW10
JH13
JH4
JH8
JH6
JH5
CW21
KG8
JH1
WS-53
77
PBP
TSC2
10kb
60kb
M
R

Fig. 6

```
   1  CTCAACGAGGAGCCCCTGACGCTGGCGGGCGAGGAGATCGTGGCCCAGGGCAAGCGCTCG   60
   1  L  N  E  E  P  L  T  L  A  G  E  E  I  V  A  Q  G  K  R  S   20

  61  GACCCGCGGAGCCTGCTGTGCTATGGCGGCGCCCCAGGGCCTGGCTGCCACTTCTCCATC  120
  21  D  P  R  S  L  L  C  Y  G  G  A  P  G  P  G  C  H  F  S  I   40

 121  CCCGAGGCTTTCAGCGGGGCCCTGGCCAACCTCAGTGACGTGGTGCAGCTCATCTTTCTG  180
  41  P  E  A  F  S  G  A  L  A  N  L  S  D  V  V  Q  L  I  F  L   60

 181  GTGGACTCCAATCCCTTTCCCTTTGGCTATATCAGCAACTACACCGTCTCCACCAAGGTG  240
  61  V  D  S  N  P  F  P  F  G  Y  I  S  N  Y  T  V  S  T  K  V   80

 241  GCCTCGATGGCATTCCAGACACAGGCCGCGCCCAGATCCCCATCGAGCGGCTGGCCTCA  300
  81  A  S  M  A  F  Q  T  Q  A  G  A  Q  I  P  I  E  R  L  A  S  100

 301  GAGCGCGCCATCACCGTGAAGGTGCCCAACAACTCGGACTGGGCTGCCCGGGGCCACCGC  360
 101  E  R  A  I  T  V  K  V  P  N  N  S  D  W  A  A  R  G  H  R  120

 361  AGCTCCGCCAACTCCGCCAACTCCGTTGTGGTCCAGCCCCAGGCCTCCGTGGGTGCTGTG  420
 121  S  S  A  N  S  A  N  S  V  V  V  Q  P  Q  A  S  V  G  A  V  140

 421  GTCACCCTGGACAGCAGCAACCCTGCGGCCGGGCTGCATCTGCAGCTCAACTATACGCTG  480
 141  V  T  L  D  S  S  N  P  A  A  G  L  H  L  Q  L  N  Y  T  L  160

 481  CTGGACGGCCACTACCTGTCTGAGGAACCTGAGCCCTACCTGGCAGTCTACCTACACTCG  540
 161  L  D  G  H  Y  L  S  E  E  P  E  P  Y  L  A  V  Y  L  H  S  180

 541  GAGCCCCGGCCCAATGAGCACAACTGCTCGGCTAGCAGGAGGATCCGCCCAGAGTCACTC  600
 181  E  P  R  P  N  E  H  N  C  S  A  S  R  R  I  R  P  E  S  L  200

 601  CAGGGTGCTGACCACCGGCCCTACACCTTCTTCATTTCCCCGGGGAGCAGAGACCCAGCG  660
 201  Q  G  A  D  H  R  P  Y  T  F  F  I  S  P  G  S  R  D  P  A  220

 661  GGGAGTTACCATCTGAACCTCTCCAGCCACTTCCGCTGGTCGGCGCTGCAGGTGTCCGTG  720
 221  G  S  Y  H  L  N  L  S  S  H  F  R  W  S  A  L  Q  V  S  W  240

 721  GGCCTGTACACGTCCCTGTGCCAGTACTTCAGCGAGGAGGACATGGTGTGGCGGACAGAG  780
 241  G  L  Y  T  S  L  C  Q  Y  F  S  E  E  D  M  V  W  R  T  E  260

 781  GGGCTGCTGCCCCTGGAGGAGACCTCGCCCCGCCAGGCCGTCTGCCTCACCCGCCACCTC  840
 261  G  L  L  P  L  E  E  T  S  P  R  Q  A  V  C  L  T  R  H  L  280

 841  ACCGCCTTCGGCGCCAGCCTCTTCGTGCCCCCAAGCCATGTCCGCTTTGTGTTTCCTGAG  900
 281  T  A  F  G  A  S  L  F  V  P  P  S  H  V  R  F  V  F  P  E  300

 901  CCGACAGCGGATGTAAACTACATCGTCATGCTGACATGTGCTGTGTGCCTGGTGACCTAC  960
 301  P  T  A  D  V  N  Y  I  V  M  L  T  C  A  V  C  L  V  T  Y  320

 961  ATGGTCATGGCCGCCATCCTGCACAAGCTGGACCAGTTGGATGCCAGCCGGGGCCGCGCC 1020
 321  M  V  M  A  A  I  L  H  K  L  D  Q  L  D  A  S  R  G  R  A  340

1021  ATCCCTTTCTGTGGGCAGCGGGGCCGCTTCAAGTACGAGATCCTCGTCAAGACAGGCTGG 1080
 341  I  P  F  C  G  Q  R  G  R  F  K  Y  E  I  L  V  K  T  G  W  360

1081  GGCCGGGGCTCAGGTACCACGGCCCACGTGGGCATCATGCTGTATGGGGTGGACAGCCGG 1140
 361  G  R  G  S  G  T  T  A  H  V  G  I  M  L  Y  G  V  D  S  R  380

1141  AGCGGCCACCGGCACCTGGACGGCGACAGAGCCTTCCACCGCAACAGCCTGGACATCTTC 1200
 381  S  G  H  R  H  L  D  G  D  R  A  F  H  R  N  S  L  D  I  F  400

1201  CGGATGGCCACCCCGCACAGCCTGGGTAGCGTGTGGAAGATCCGAGTGTGGCACGACAAC 1260
 401  R  I  A  T  P  H  S  L  G  S  V  W  K  I  R  V  W  H  D  N  420
```

Figure 7

```
1261 AAAGGGCTCAGCCCTGCCTGGTTCCTGCAGCACGTCATCGTCAGGGACCTGCAGACGGCA 1320
 421 K  G  L  S  P  A  W  F  L  Q  H  V  I  V  R  D  L  Q  T  A   440

1321 CGCAGCGCCTTCTTCCTGGTCAATGACTGGCTTTCGGTGGAGACGGAGGCCAACGGGGGC 1380
 441 R  S  A  F  F  L  V  N  D  W  L  S  V  E  T  E  A  N  G  G   460

1381 CTGGTGGAGAAGGAGGTGCTGGCCGCGAGCGACGCAGCCCTTTTGCGCTTCCGGCGCCTG  140
 461 L  V  E  K  E  V  L  A  A  S  D  A  A  L  L  R  F  R  R  L   480

1441 CTGGTGGCTGAGCTGCAGCGTGGCTTCTTTGACAAGCACATCTGGCTCTCCATATGGGAC 1500
 481 L  V  A  E  L  Q  R  G  F  F  D  K  H  I  W  L  S  I  W  D   500

1501 CGGCCGCCTCGTAGCCGTTTCACTGCATCCAGAGGGCCACCTGCTGCGTTCTCCTCATC 1560
 501 R  P  P  R  S  R  F  T  R  I  Q  R  A  T  C  C  V  L  L  I   520

1561 TGCCTCTTCCTGGGCGCCAACGCCGTGTGGTACGGGGCTGTTGGCGACTCTGCCTACAGC 1620
 521 C  L  F  L  G  A  N  A  V  W  Y  G  A  V  G  D  S  A  Y  S   540

1621 ACGGGGCATGTGTCCAGGCTGAGCCCGCTGAGCGTCGACACAGTCGCTGTTGGCCTCGTG 1680
 541 T  G  H  V  S  R  L  S  P  L  S  V  D  T  V  A  V  G  L  V   560

1681 TCCAGCGTGGTTGTCTATCCCGTCTACCTGGCCATCCTTTTTCTCTTCCGGATGTCCCGG 1740
 561 S  S  V  V  V  Y  P  V  Y  L  A  I  L  F  L  F  R  M  S  R   580

1741 AGCAAGGTGGCTGGGAGCCCGAGCCCCCACACTGCCGGGCAGCAGGTGCTGGACATCGAC 1800
 581 S  K  V  A  G  S  P  S  P  T  P  A  G  Q  Q  V  L  D  I  D   600

1801 AGCTGCCTGGACTCGTCCGTGCTGGACAGCTCCTTCCTCACGTTCTCAGGCCTCCACGCT 1860
 601 S  C  L  D  S  S  V  L  D  S  S  F  L  T  F  S  G  L  H  A   620

1861 GAGGCCTTTGTTGGACAGATGAAGAGTGACTTGTTTCTGGATGATTCTAAGAGTCTGGTG 1920
 621 E  A  F  V  G  Q  M  K  S  D  L  F  L  D  D  S  K  S  L  V   640

1921 TGCTGGCCCTCCGGCGAGGGAACGCTCAGTTGGCCGCACCTGCTCAGTGACCCGTCCATT 1980
 641 C  W  P  S  G  E  G  T  L  S  W  P  D  L  L  S  D  P  S  I   660

1981 GTGGGTAGCAATCTGCGGCAGCTGGCACGGGGCCAGGCGGGCCATGGGCTGGGCCCAGAG 2040
 661 V  G  S  N  L  R  Q  L  A  R  G  Q  A  G  H  G  L  G  P  E   680

2041 GAGGACGGCTTCTCCCTGGCCAGCCCCTCCTGCCTGCCAAATCCTTCTCAGCATCAGAT 2100
 681 E  D  G  F  S  L  A  S  P  Y  S  P  A  K  S  F  S  A  S  D   700

2101 GAAGACCTGATCCAGGAGGTCCTTGCCGAGGGGGTCAGCAGCCCAGCCCCTACCCAAGAC 2160
 701 E  D  L  I  Q  Q  V  L  A  E  G  V  S  S  P  A  P  T  Q  D   720

2161 ACCCACATGGAAACGGACCTGCTCAGCAGCCTGTCCAGCACTCCTGGGGAGAAGACAGAG 2220
 721 T  H  M  E  T  D  L  L  S  S  L  S  S  T  P  G  E  K  T  E   740

2221 ACGCTGGCGCTGCAGAGGCTGGGGGAGCTGGGGCCACCCAGCCCAGGCCTGAACTGGGAA 2280
 741 T  L  A  L  Q  R  L  G  E  L  G  P  P  S  P  G  L  N  W  E   760

2281 CAGCCCCAGGCAGCGAGGCTGTCCAGGACAGGACTGGTGGAGGGTCTGCGGAAGCGCCTG 2340
 761 Q  P  Q  A  A  R  L  S  R  T  G  L  V  E  G  L  R  K  R  L   780

2341 CTGCCGGCCTGGTGTGCCTCCCTGGCCCACGGGCTCAGCCTGCTCCTGGTGGCTGTGGCT 2400
 781 L  P  A  W  C  A  S  L  A  H  G  L  S  L  L  L  V  A  V  A   800

2401 GTGGCTGTCTCAGGGTGGGTGGGTGCGAGCTTCCCCCCGGGCCTGAGTGTTGCGTGGCTC 2460
 801 V  A  V  S  G  W  V  G  A  S  F  P  P  G  V  S  V  A  W  L   820

2461 CTGTCCAGCAGCGCCAGCTTCCTGGCCTCATTCCTCGGCTGGGAGCCACTGAAGGTCTTG 2520
 821 L  S  S  S  A  S  F  L  A  S  F  L  G  W  E  P  L  K  V  L   840
```

Figure 7 cont'd

```
2521  CTGGAAGCCCTGTACTTCTCACTGGTGGCCAAGCGGCTGCACCCGGATGAAGATGACACC  2580
 841  L  E  A  L  Y  F  S  L  V  A  K  R  L  H  P  D  E  D  D  T    860

2581  CTGGTAGAGAGCCCCGCTGTGACGCCTGTGAGCGCACGTGTGCCCCGCGTACGGCCACCC  2640
 861  L  V  E  S  P  A  V  T  P  V  S  A  R  V  P  R  V  R  P  P    880

2641  CACGGCTTTGCACTCTTCCTGGCCAAGGAAGAAGCCCGCAAGGTCAAGAGGCTACATGGC  2700
 881  H  G  F  A  L  F  L  A  K  E  E  A  R  K  V  K  R  L  H  G    900

2701  ATGCTGCGGAGCCTCCTGGTGTACATGCTTTTTCTGCTGGTGACCCTGCTGGCCAGCTAT  2760
 901  M  L  R  S  L  L  V  Y  M  L  F  L  L  V  T  L  L  A  S  Y    920

2761  GGGGATGCCTCATGCCATGGGCACGCCTACGGTCTGCAAAGCGCCATCAAGCAGGAGCTG  2820
 921  G  D  A  S  C  H  G  H  A  Y  R  L  Q  S  A  I  K  Q  E  L    940

2821  CACAGCCGGGCCTTCCTGGCCATCACGCGGTCTGAGGAGCTCTGGCCATGGATGGCCCAC  2880
 941  H  S  R  A  F  L  A  I  T  R  S  E  E  L  W  P  W  M  A  H    960

2881  GTGCTGCTGCCCTACGTCCACGGGAACCAGTCCAGCCCAGAGCTGGGGCCCCCACGGCTG  2940
 961  V  L  L  P  Y  V  H  G  N  Q  S  S  P  E  L  G  P  P  R  L    980

2941  CGGCAGGTGCGGCTGCAGGAAGCACTCTACCCAGACCCTCCCGGCCCC'AGGGTCCACACG  3000
 981  R  Q  V  R  L  Q  E  A  L  Y  P  D  P  P  G  P  R  V  H  T   1000

3001  TGCTCGGCCGCAGGAGGCTTCAGCACCAGCGATTACGACGTTGGCTGGGAGAGTCCTCAC  3060
1001  C  S  A  A  G  G  F  S  T  S  D  Y  D  V  G  W  E  S  P  H   1020

3061  AATGGCTCGGGGACGTGGGCCTATTCAGCGCCGGATCTGCTGGGGGCATGGTCCTGGGGC  3120
1021  N  G  S  G  T  W  A  Y  S  A  P  D  L  L  G  A  W  S  W  G   1040

3121  TCCTGTGCCGTGTATGACAGCGGGGGCTACGTGCAGGAGCTGGGCCTGAGCCTGGAGGAG  3180
1041  S  C  A  V  Y  D  S  G  G  Y  V  Q  E  L  G  L  S  L  E  E   1060

3181  AGCCGCGACCGGCTGCGCTTCCTGCAGCTGCACAACTGGCTGGACAACAGGAGCCGCGCT  3240
1061  S  R  D  R  L  R  F  L  Q  L  H  N  W  L  D  N  R  S  R  A   1080

3241  GTGTTCCTGGAGCTCACGCGCTACAGCCCGGCCGTGGGGCTGCACGCCGCCGTCACGCTG  3300
1081  V  F  L  E  L  T  R  Y  S  P  A  V  G  L  H  A  A  V  T  L   1100

3301  CGCCTCGAGTTCCCGGGCGCCGGCCGCGCCCTGGCCGCCCTCAGCGTCCGCCCCTTTGCG  3360
1101  R  L  E  F  P  A  A  G  R  A  L  A  A  L  S  V  R  P  F  A   1120

3361  CTGCGCCGCCTCAGCGCGGGCCTCTCGCTGCCTCTGCTCACCTCGGTGTGCCTGCTGCTG  3420
1121  L  R  R  L  S  A  G  L  S  L  P  L  L  T  S  V  C  L  L  L   1140

3421  TTCGCCGTGCACTTCGCCGTGGCCGAGGCCCGTACTTGGCACAGGGAAGGGGCTGGCGC  3480
1141  F  A  V  H  F  A  V  A  E  A  R  T  W  H  R  E  G  R  W  R   1160

3481  GTGCTGCGGCTCGGAGCCTGGGCGCGGTGGCTGCTGGTGGCGCTGACGGCGGCCACGGCA  3540
1161  V  L  R  L  G  A  W  A  R  W  L  L  V  A  L  T  A  A  T  A   1180

3541  CTGGTACGCCTCGCCCAGCTGGGTGCCGCTGACCGCCAGTGGACCCGTTTCGTGCGCGGC  3600
1181  L  V  R  L  A  Q  L  G  A  A  D  R  Q  W  T  R  F  V  R  G   1200

3601  CGCCCGCGCCGCTTCACTAGCTTCGACCAGGTGGCGCACGTGAGCTCCGCAGCCCGTGGC  3660
1201  R  P  R  R  F  T  S  F  D  Q  V  A  H  V  S  S  A  A  R  G   1220

3661  CTGGCGGCCTCGCTGCTCTTCCTGCTTTTGGTCAAGGCTGCCCAGCACGTACGCTTCGTG  3720
1221  L  A  A  S  L  L  F  L  L  L  V  K  A  A  Q  H  V  R  F  V   1240

3721  CGCCAGTGGTCCGTCTTTGGCAAGACATTATGCCGAGCTCTGCCAGAGCTCCTGGGGGTC  3780
1241  R  Q  W  S  V  F  G  K  T  L  C  R  A  L  P  E  L  L  G  V   1260
```

Figure 7 Cont'd

```
5101  TAGGGCTGAGGGGCCTGCGCCCAGAGCTGGCCTCCCCCAACACCTGCTGCGCTTGGTAGG  5160
5161  TGTGGTGGCGTTATGGCAGCCCGGCTGCTGCTTGGATGCGAGCTTGGCCTTGGGCCGGTG  5220
5221  CTGGGGGCACAGCTGTCTGCCAGGCACTCTCATCACCCCAGAGGCCTTGTCATCCTCCCT  5280
5281  TGCCCCAGGCCAGGTAGCAAGAGAGCAGCGCCCAGGCCTGCTGGCATCAGGTCTGGGCAA  5340
5341  CTAGCAGGACTAGGCATGTCAGAGGACCCCAGGGTGGTTAGAGGAAAAGACTCCTCCTGG  5400
5401  GGGCTGGCTCCCAGGGTGGAGGAAGGTGACTGTGTGTGTGTGTGTGTGCGCGCGCGACGC  5460
5461  GCGACTGTGCTGTATGGCCCAGGCACGCTCAAGGCCCTCGGAGCTGGCTGTGCCTGCTTC  5520
5521  TGTGTACCACTTCTGTGGGCATGGCCGCTTCTAGAGCCTCGACACCCCCCCAACCCCCGC  5580
5581  ACCAAGCAGACAAAGTCAATAAAAGAGCTGTCTGACTGCAAAAAAAAAAAA 5631
```

1A1H0.6

```
   1  AAGCTTGGCA CCATCAAGGG CCAGTTCAAC TTTGTCCACG TGATCGTCAC CCCGCTGGAC
  61  TACGAGTGCA ACCTGGTGTC CCTGCAGTGC AGGAAAGACA TGGAGGGCCT TGTGGACACC
 121  AGCGTGGCCA AGATCGTGTC TGACCGCAAC CTGCCCTTCG TGGCCGCCA GATGGCCCTG
 181  CACGCAAATA TGGCCTCACA GGTGCATCAT AGCCGCTCCA ACCCCACCGA TATCTACCCC
 241  TCCAAGTGGA TTGCCGGCT CCGCCACATC AAGCGGCTCC GCCAGCGGAT CTGCGAGGAA
 301  GCCGCCTACT CCAACCCCAG CCTACCTCTG GTGCACCCTC CGTCCCATAG CAAAGCCCCT
 361  GCACAGACTC CAGCCGAGCC CACACCTGGC TATGAGGTGG GCCAGCGGAA GCGCCTCATC
 421  TCCTCGGTGG AGGACTTCAC CGAGTTTGTG TGAGGCCGGG GCCCTCCCTC CTGCACTGGC
 481  CTTGGACGGT ATTGCCTGTC AGTGAAATAA ATAAAGTCCT GACCCCAGTG CACAGACATA
 541  GAGGCACAGA TTGC
```

Figure 8

CW10L

```
   1  GTCCGCGGTC GCACGTACGC TTCTGGTGTG TGTGAGACGT GCGGGGCTGG GAAGTGTTGG
  61  CAGACGGCGA GTACGTCCTC ACTCCTTTTG TTCTTTTGAC CTAAGCTGGC GAGTGGCACT
 121  GCTGAGTTCC GCTCAGTGCC CGCCCTGATG TGCGACCCCC GTGCATTCTT GCTGTTAGGT
 181  GGTGGCGGTG TG
```

CW10R

```
   1  AGGCAGGTCT CCCCCACGAG CAGGGGAGAG GCACCCAAGG T.
```

Figure 9

Figure 7 cont'd


```
3781  ACCTTGGGCCTGGTGGTGCTCGGGGTAGCCTACGCCCAGCTGGCCATCCTGCTCGTGTCT  3840
1261  T  L  G  L  V  V  L  G  V  A  Y  A  Q  L  A  I  L  L  V  S   1280

3841  TCCTGTGTGGACTCCCTCTGGAGCGTGGCCCAGGCCCTGTTGGTGCTGTGCCCTGGGACT  3900
1281  S  C  V  D  S  L  W  S  V  A  Q  A  L  L  V  L  C  P  G  T   1300

3901  GGGCTCTCTACCCTGTGTCCTGCCGAGTCCTGGCACCTGTCACCCCTGCTGTGTGTGGGG  3960
1301  G  L  S  T  L  C  P  A  E  S  W  H  L  S  P  L  L  C  V  G   1320

3961  CTCTGGGCACTGCGGCTGTGGGGCGCCCTACGGCTGGGGGCTGTTATTCTCCGCTGGCGC  4020
1321  L  W  A  L  R  L  W  G  A  L  R  L  G  A  V  I  L  R  W  R   1340

4021  TACCACGCCTTGCGTGGAGAGCTGTACCGGCCGGCCTGGGAGCCCCAGGACTACGAGATG  4080
1341  Y  H  A  L  R  G  E  L  Y  R  P  A  W  E  P  Q  D  Y  E  M   1360

4081  GTGGAGTTGTTCCTGCGCAGGCTGCGCCTCTGGATGGGCCTCAGCAAGGTCAAGGAGTTC  4140
1361  V  E  L  F  L  R  R  L  R  L  W  M  G  L  S  K  V  K  E  F   1380

4141  CGCCACAAAGTCCGCTTTGAAGGGATGGAGCCGCTGCCCTCTGCTCCTCCAGGGGCTCC  4200
1381  R  H  K  V  R  F  E  G  M  E  P  L  P  S  R  S  S  R  G  S   1400

4201  AAGGTATCCCCGGATGTGCCCCCACCCAGCGCTGGCTCCGATGCCTCGCACCCCTCCACC  4260
1401  K  V  S  P  D  V  P  P  P  S  A  G  S  D  A  S  H  P  S  T   1420

4261  TCCTCCAGCCAGCTGGATGGGCTGAGCGTGAGCCTGGGCCGGCTGGGGACAAGGTGTGAG  4320
1421  S  S  S  Q  L  D  G  L  S  V  S  L  G  R  L  G  T  R  C  E   1440

4321  CCTGAGCCCTCCCGCCTCCAAGCCGTGTTCGAGGCCCTGCTCACCCAGTTTGACCGACTC  4380
1441  P  E  P  S  R  L  Q  A  V  F  E  A  L  L  T  Q  F  D  R  L   1460

4381  AACCAGGCCACAGAGGACGTCTACCAGCTGGAGCAGCAGCTGCACAGCCTGCAAGGCCGC  4440
1461  N  Q  A  T  E  D  V  Y  Q  L  E  Q  Q  L  H  S  L  Q  G  R   1480

4441  AGGAGCAGCCGGGCGCCCGCCGGATCTTCCCGTGGCCCATCCCCGGGCCTGCGGCCAGCA  4500
1481  R  S  S  R  A  P  A  G  S  S  R  G  P  S  P  G  L  R  P  A   1500

4501  CTGCCCAGCCGCCTTGCCCGGGCCAGTCGGGGTGTGGACCTGGCCACTGGCCCCAGCAGG  4560
1501  L  P  S  R  L  A  R  A  S  R  G  V  D  L  A  T  G  P  S  R   1520

4561  ACACCTTCGGGCCAAGAACAAGGTCCACCCCAGCAGCACTTAGTCCTCCTTCCTGGCGGG  4620
1521  T  P  S  G  Q  E  Q  G  P  P  Q  Q  H  L  V  L  L  P  G  G   1540

4621  GGTGGGCCGTGGAGTCGGAGTGGACACCGCTCAGTATTACTTTCTGCCGCTGTCAAGGCC  46890
1541  G  G  P  W  S  R  S  G  H  R  S  V  L  L  S  A  A  V  K  A   1560

4681  GAGGGCCAGGCAGAATGGCTGCACGTAGGTTCCCCAGAGAGCAGGCAGGGGCATCTGTCT  4740
1561  E  G  Q  A  E  W  L  H  V  G  S  P  E  S  R  Q  G  H  L  S   1580

4741  GTCTGTGGGCTTCAGCACTTTAAAGAGGCTGTGTGGCCAACCAGGACCCAGGGTCCCCTC  4800
1581  V  C  G  L  Q  H  F  K  E  A  V  W  P  T  R  T  Q  G  P  L   1600

4801  CCCAGCTCCCTTGGGAAGGACACAGCAGTATTGGACGGTTTCTAGCCTCTGAGATGCTAA  4860
1601  P  S  S  L  G  K  D  T  A  V  L  D  G  F                     1620

4861  TTTATTTCCCCGAGTCCTCAGGTACAGCGGGCTGTGCCCGGCCCCACCCCCTGGGCAGAT  4920

4921  GTCCCCCACTGCTAAGGCTGCTGGCTTCAGGGAGGGTTAGCCTGCACCGCCGCCACCCTG  4980

4981  CCCCTAAGTTATTACCTCTCCAGTTCCTACCGTACTCCCTGCACCGTCTCACTGTGTGTC  5040

5041  TCGTCTCAGTAATTTATATGGTGTTAAAATGTGTATATTTTTGTATGTCACTATTTTCAC  5100
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1: (Compare Fig.1)

```
C GGC GCC GCC TGC CGC GTC AAC TGC TCG GGC CGC GGG CTG CGG ACG      46
  Gly Ala Ala Cys Arg Val Asn Cys Ser Gly Arg Gly Leu Arg Thr
    1               5                  10                  15

CTC GGT CCC GCG CTG CGC ATC CCC GCG GAC GCC ACA GCG CTA GAC GTC      94
Leu Gly Pro Ala Leu Arg Ile Pro Ala Asp Ala Thr Ala Leu Asp Val
                20                  25                  30

TCC CAC AAC CTG CTC CGG GCG CTG GAC GTT GGG CTC CTG GCG AAC CTC     142
Ser His Asn Leu Leu Arg Ala Leu Asp Val Gly Leu Leu Ala Asn Leu
            35                  40                  45

TCG GCG CTG GCA GAG CTG GAT ATA AGC AAC AAC AAG ATT TCT ACG TTA     190
Ser Ala Leu Ala Glu Leu Asp Ile Ser Asn Asn Lys Ile Ser Thr Leu
        50                  55                  60

GAA GAA GGA ATA TTT GCT AAT TTA TTT AAT TTA AGT GAA ATA AAC CTG     238
Glu Glu Gly Ile Phe Ala Asn Leu Phe Asn Leu Ser Glu Ile Asn Leu
    65                  70                  75

AGT GGG AAC CCG TTT GAG TGT GAC TGT GGC CTG GCG TGG CTG CCG CGA     286
Ser Gly Asn Pro Phe Glu Cys Asp Cys Gly Leu Ala Trp Leu Pro Arg
80                  85                  90                  95

TGG GCG GAG GAG CAG CAG GTG CGG GTG GTG CAG CCC GAG GCA GCC ACG     334
Trp Ala Glu Glu Gln Gln Val Arg Val Val Gln Pro Glu Ala Ala Thr
                100                 105                 110

TGT GCT GGG CCT GGC TCC CTG GCT GGC CAG CCT CTG CTT GGC ATC CCC     382
Cys Ala Gly Pro Gly Ser Leu Ala Gly Gln Pro Leu Leu Gly Ile Pro
            115                 120                 125

TTG CTG GAC AGT GGC TGT GGT GAG GAG TAT GTC GCC TGC CTC CCT GAC     430
Leu Leu Asp Ser Gly Cys Gly Glu Glu Tyr Val Ala Cys Leu Pro Asp
        130                 135                 140

AAC AGC TCA GGC ACC GTG GCA GCA GTG TCC TTT TCA GCT GCC CAC GAA     478
Asn Ser Ser Gly Thr Val Ala Ala Val Ser Phe Ser Ala Ala His Glu
    145                 150                 155

GGC CTG CTT CAG CCA GAG GCC TGC AGC GCC TTC TGC TTC TCC ACC GGC     526
Gly Leu Leu Gln Pro Glu Ala Cys Ser Ala Phe Cys Phe Ser Thr Gly
160                 165                 170                 175

CAG GGC CTC GCA GCC CTC TCG GAG CAG GGC TGG TGC CTG TGT GGG GCG     574
Gln Gly Leu Ala Ala Leu Ser Glu Gln Gly Trp Cys Leu Cys Gly Ala
                180                 185                 190

GCC CAG CCC TCC AGT GCC TCC TTT GCC TGC CTG TCC CTC TGC TCC GGC     622
Ala Gln Pro Ser Ser Ala Ser Phe Ala Cys Leu Ser Leu Cys Ser Gly
            195                 200                 205

CCC CCG CCA CCT CCT GCC CCC ACC TGT AGG GGC CCC ACC CTC CTC CAG     670
Pro Pro Pro Pro Pro Ala Pro Thr Cys Arg Gly Pro Thr Leu Leu Gln
        210                 215                 220

CAC GTC TTC CCT GCC TCC CCA GGG GCC ACC CTG GTG GGG CCC CAC GGA     718
His Val Phe Pro Ala Ser Pro Gly Ala Thr Leu Val Gly Pro His Gly
    225                 230                 235
```

```
CCT CTG GCC TCT GGC CAG CTA GCA GCC TTC CAC ATC GCT GCC CCG CTC      766
Pro Leu Ala Ser Gly Gln Leu Ala Ala Phe His Ile Ala Ala Pro Leu
240                 245                 250                 255

CCT GTC ACT GCC ACA CGC TGG GAC TTC GGA GAC GGC TCC GCC GAG GTG      814
Pro Val Thr Ala Thr Arg Trp Asp Phe Gly Asp Gly Ser Ala Glu Val
                260                 265                 270

GAT GCC GCT GGG CCG GCT GCC TCG CAT CGC TAT GTG CTG CCT GGG CGC      862
Asp Ala Ala Gly Pro Ala Ala Ser His Arg Tyr Val Leu Pro Gly Arg
            275                 280                 285

TAT CAC GTG ACG GCC GTG CTG GCC CTG GGG GCC GGC TCA GCC CTG CTG      910
Tyr His Val Thr Ala Val Leu Ala Leu Gly Ala Gly Ser Ala Leu Leu
        290                 295                 300

GGG ACA GAC GTG CAG GTG GAA GCG GCA CCT GCC GCC CTG GAG CTC GTG      958
Gly Thr Asp Val Gln Val Glu Ala Ala Pro Ala Ala Leu Glu Leu Val
    305                 310                 315

TGC CCG TCC TCG GTG CAG AGT GAC GAG AGC CTT GAC CTC AGC ATC CAG     1006
Cys Pro Ser Ser Val Gln Ser Asp Glu Ser Leu Asp Leu Ser Ile Gln
320                 325                 330                 335

AAC CGC GGT GGT TCA GGC CTG GAG GCC GCC TAC AGC ATC GTG GCC CTG     1054
Asn Arg Gly Gly Ser Gly Leu Glu Ala Ala Tyr Ser Ile Val Ala Leu
                340                 345                 350

GGC GAG GAG CCG GCC CGA GCG GTG CAC CCG CTC TGC CCC TCG GAC ACG     1102
Gly Glu Glu Pro Ala Arg Ala Val His Pro Leu Cys Pro Ser Asp Thr
            355                 360                 365

GAG ATC TTC CCT GGC AAC GGG CAC TGC TAC CGC CTG GTG GTG GAG AAG     1150
Glu Ile Phe Pro Gly Asn Gly His Cys Tyr Arg Leu Val Val Glu Lys
        370                 375                 380

GCG GCC TGG CTG CAG GCG CAG GAG CAG TGT CAG GCC TGG GCC GGG GCC     1198
Ala Ala Trp Leu Gln Ala Gln Glu Gln Cys Gln Ala Trp Ala Gly Ala
    385                 390                 395

GCC CTG GCA ATG GTG GAC AGT CCC GCC GTG CAG CGC TTC CTG GTC TCC     1246
Ala Leu Ala Met Val Asp Ser Pro Ala Val Gln Arg Phe Leu Val Ser
400                 405                 410                 415

CGG GTC ACC AGG AGC CTA GAC GTG TGG ATC GGC TTC TCG ACT GTG CAG     1294
Arg Val Thr Arg Ser Leu Asp Val Trp Ile Gly Phe Ser Thr Val Gln
                420                 425                 430

GGG GTG GAG GTG GGC CCA GCG CCG CAG GGC GAG GCC TTC AGC CTG GAG     1342
Gly Val Glu Val Gly Pro Ala Pro Gln Gly Glu Ala Phe Ser Leu Glu
            435                 440                 445

AGC TGC CAG AAC TGG CTG CCC GGG GAG CCA CAC CCA GCC ACA GCC GAG     1390
Ser Cys Gln Asn Trp Leu Pro Gly Glu Pro His Pro Ala Thr Ala Glu
        450                 455                 460

CAC TGC GTC CGG CTC GGG CCC ACC GGG TGG TGT AAC ACC GAC CTG TGC     1438
His Cys Val Arg Leu Gly Pro Thr Gly Trp Cys Asn Thr Asp Leu Cys
    465                 470                 475
```

```
TCA GCG CCG CAC AGC TAC GTC TGC GAG CTG CAG CCC GGA GGC CCA GTG      1486
Ser Ala Pro His Ser Tyr Val Cys Glu Leu Gln Pro Gly Gly Pro Val
480                 485                 490                 495

CAG GAT GCC GAG AAC CTC CTC GTG GGA GCG CCC AGT GGG GAC CTG CAG      1534
Gln Asp Ala Glu Asn Leu Leu Val Gly Ala Pro Ser Gly Asp Leu Gln
                    500                 505                 510

GGA CCC CTG ACG CCT CTG GCA CAG CAG GAC GGC CTC TCA GCC CCG CAC      1582
Gly Pro Leu Thr Pro Leu Ala Gln Gln Asp Gly Leu Ser Ala Pro His
            515                 520                 525

GAG CCC GTG GAG GTC ATG GTA TTC CCG GGC CTG CGT CTG AGC CGT GAA      1630
Glu Pro Val Glu Val Met Val Phe Pro Gly Leu Arg Leu Ser Arg Glu
        530                 535                 540

GCC TTC CTC ACC ACG GCC GAA TTT GGG ACC CAG GAG CTC CGG CGG CCC      1678
Ala Phe Leu Thr Thr Ala Glu Phe Gly Thr Gln Glu Leu Arg Arg Pro
    545                 550                 555

GCC CAG CTG CGG CTG CAG GTG TAC CGG CTC CTC AGC ACA GCA GGG ACC      1726
Ala Gln Leu Arg Leu Gln Val Tyr Arg Leu Leu Ser Thr Ala Gly Thr
560                 565                 570                 575

CCG GAG AAC GGC AGC GAG CCT GAG AGC AGG TCC CCG GAC AAC AGG ACC      1774
Pro Glu Asn Gly Ser Glu Pro Glu Ser Arg Ser Pro Asp Asn Arg Thr
                580                 585                 590

CAG CTG GCC CCC GCG TGC ATG CCA GGG GGA CGC TGG TGC CCT GGA GCC      1822
Gln Leu Ala Pro Ala Cys Met Pro Gly Gly Arg Trp Cys Pro Gly Ala
            595                 600                 605

AAC ATC TGC TTG CCG CTG GAC GCC TCT TGC CAC CCC CAG GCC TGC GCC      1870
Asn Ile Cys Leu Pro Leu Asp Ala Ser Cys His Pro Gln Ala Cys Ala
        610                 615                 620

AAT GGC TGC ACG TCA GGG CCA GGG CTA CCC GGG GCC CCC TAT GCG CTA      1918
Asn Gly Cys Thr Ser Gly Pro Gly Leu Pro Gly Ala Pro Tyr Ala Leu
    625                 630                 635

TGG AGA GAG TTC CTC TTC TCC GTT GCC GCG GGG CCC CCC GCG CAG TAC      1966
Trp Arg Glu Phe Leu Phe Ser Val Ala Ala Gly Pro Pro Ala Gln Tyr
640                 645                 650                 655

TCG GTC ACC CTC CAC GGC CAG GAT GTC CTC ATG CTC CCT GGT GAC CTC      2014
Ser Val Thr Leu His Gly Gln Asp Val Leu Met Leu Pro Gly Asp Leu
                660                 665                 670

GTT GGC TTG CAG CAC GAC GCT GGC CCT GGC GCC CTC CTG CAC TGC TCG      2062
Val Gly Leu Gln His Asp Ala Gly Pro Gly Ala Leu Leu His Cys Ser
            675                 680                 685

CCG GCT CCC GGC CAC CCT GGT CCC CAG GCC CCG TAC CTC TCC GCC AAC      2110
Pro Ala Pro Gly His Pro Gly Pro Gln Ala Pro Tyr Leu Ser Ala Asn
        690                 695                 700

GCC TCG TCA TGG CTG CCC CAC TTG CCA GCC CAG CTG GAG GGC ACT TGG      2158
Ala Ser Ser Trp Leu Pro His Leu Pro Ala Gln Leu Glu Gly Thr Trp
    705                 710                 715
```

```
GCC TGC CCT GCC TGT GCC CTG CGG CTG CTT GCA GCC ACG GAA CAG CTC      2206
Ala Cys Pro Ala Cys Ala Leu Arg Leu Leu Ala Ala Thr Glu Gln Leu
720                 725                 730                 735

ACC GTG CTG CTG GGC TTG AGG CCC AAC CCT GGA CTG CGG ATG CCT GGG      2254
Thr Val Leu Leu Gly Leu Arg Pro Asn Pro Gly Leu Arg Met Pro Gly
                740                 745                 750

CGC TAT GAG GTC CGG GCA GAG GTG GGC AAT GGC GTG TCC AGG CAC AAC      2302
Arg Tyr Glu Val Arg Ala Glu Val Gly Asn Gly Val Ser Arg His Asn
            755                 760                 765

CTC TCC TGC AGC TTT GAC GTG GTC TCC CCA GTG GCT GGG CTG CGG GTC      2350
Leu Ser Cys Ser Phe Asp Val Val Ser Pro Val Ala Gly Leu Arg Val
        770                 775                 780

ATC TAC CCT GCC CCC CGC GAC GGC CGC CTC TAC GTG CCC ACC AAC GGC      2398
Ile Tyr Pro Ala Pro Arg Asp Gly Arg Leu Tyr Val Pro Thr Asn Gly
    785                 790                 795

TCA GCC TTG GTG CTC CAG GTG GAC TCT GGT GCC AAC GCC ACG GCC ACG      2446
Ser Ala Leu Val Leu Gln Val Asp Ser Gly Ala Asn Ala Thr Ala Thr
800                 805                 810                 815

GCT CGC TGG CCT GGG GGC AGT GTC AGC GCC CGC TTT GAG AAT GTC TGC      2494
Ala Arg Trp Pro Gly Gly Ser Val Ser Ala Arg Phe Glu Asn Val Cys
                820                 825                 830

CCT GCC CTG GTG GCC ACC TTC GTG CCC GGC TGC CCC TGG GAG ACC AAC      2542
Pro Ala Leu Val Ala Thr Phe Val Pro Gly Cys Pro Trp Glu Thr Asn
            835                 840                 845

GAT ACC CTG TTC TCA GTG GTA GCA CTG CCG TGG CTC AGT GAG GGG GAG      2590
Asp Thr Leu Phe Ser Val Val Ala Leu Pro Trp Leu Ser Glu Gly Glu
        850                 855                 860

CAC GTG GTG GAC GTG GTG GTG GAA AAC AGC GCC AGC CGG GCC AAC CTC      2638
His Val Val Asp Val Val Val Glu Asn Ser Ala Ser Arg Ala Asn Leu
    865                 870                 875

AGC CTG CGG GTG ACG GCG GAG GAG CCC ATC TGT GGC CTC CGC GCC ACG      2686
Ser Leu Arg Val Thr Ala Glu Glu Pro Ile Cys Gly Leu Arg Ala Thr
880                 885                 890                 895

CCC AGC CCC GAG GCC CGT GTA CTG CAG GGA GTC CTA GTG AGG TAC AGC      2734
Pro Ser Pro Glu Ala Arg Val Leu Gln Gly Val Leu Val Arg Tyr Ser
                900                 905                 910

CCC GTG GTG GAG GCC GGC TCG GAC ATG GTC TTC CGG TGG ACC ATC AAC      2782
Pro Val Val Glu Ala Gly Ser Asp Met Val Phe Arg Trp Thr Ile Asn
            915                 920                 925

GAC AAG CAG TCC CTG ACC TTC CAG AAC GTG GTC TTC AAT GTC ATT TAT      2830
Asp Lys Gln Ser Leu Thr Phe Gln Asn Val Val Phe Asn Val Ile Tyr
        930                 935                 940

CAG AGC GCG GCG GTC TTC AAG CTC TCA CTG ACG GCC TCC AAC CAC GTG      2878
Gln Ser Ala Ala Val Phe Lys Leu Ser Leu Thr Ala Ser Asn His Val
    945                 950                 955
```

```
AGC AAC GTC ACC GTG AAC TAC AAC GTA ACC GTG GAG CGG ATG AAC AGG      2926
Ser Asn Val Thr Val Asn Tyr Asn Val Thr Val Glu Arg Met Asn Arg
960                 965                 970                 975

ATG CAG GGT CTG CAG GTC TCC ACA GTG CCG GCC GTG CTG TCC CCC AAT      2974
Met Gln Gly Leu Gln Val Ser Thr Val Pro Ala Val Leu Ser Pro Asn
                980                 985                 990

GCC ACA CTG GTA CTG ACG GGT GGT GTG CTG GTG GAC TCA GCT GTG GAG      3022
Ala Thr Leu Val Leu Thr Gly Gly Val Leu Val Asp Ser Ala Val Glu
            995                 1000                1005

GTG GCC TTC CTG TGG AAC TTT GGG GAT GGG GAG CAG GCC CTC CAC CAG      3070
Val Ala Phe Leu Trp Asn Phe Gly Asp Gly Glu Gln Ala Leu His Gln
        1010                1015                1020

TTC CAG CCT CCG TAC AAC GAG TCC TTC CCG GTT CCA GAC CCC TCG GTG      3118
Phe Gln Pro Pro Tyr Asn Glu Ser Phe Pro Val Pro Asp Pro Ser Val
    1025                1030                1035

GCC CAG GTG CTG GTG GAG CAC AAT GTC ATG CAC ACC TAC GCT GCC CCA      3166
Ala Gln Val Leu Val Glu His Asn Val Met His Thr Tyr Ala Ala Pro
1040                1045                1050                1055

GGT GAG TAC CTC CTG ACC GTG CTG GCA TCT AAT GCC TTC GAG AAC CTG      3214
Gly Glu Tyr Leu Leu Thr Val Leu Ala Ser Asn Ala Phe Glu Asn Leu
                1060                1065                1070

ACG CAG CAG GTG CCT GTG AGC GTG CGC GCC TCC CTG CCC TCC GTG GCT      3262
Thr Gln Gln Val Pro Val Ser Val Arg Ala Ser Leu Pro Ser Val Ala
            1075                1080                1085

GTG GGT GTG AGT GAC GGC GTC CTG GTG GCC GGC CGG CCC GTC ACC TTC      3310
Val Gly Val Ser Asp Gly Val Leu Val Ala Gly Arg Pro Val Thr Phe
        1090                1095                1100

TAC CCG CAC CCG CTG CCC TCG CCT GGG GGT GTT CTT TAC ACG TGG GAC      3358
Tyr Pro His Pro Leu Pro Ser Pro Gly Gly Val Leu Tyr Thr Trp Asp
    1105                1110                1115

TTC GGG GAC GGC TCC CCT GTC CTG ACC CAG AGC CAG CCG GCT GCC AAC      3406
Phe Gly Asp Gly Ser Pro Val Leu Thr Gln Ser Gln Pro Ala Ala Asn
1120                1125                1130                1135

CAC ACC TAT GCC TCG AGG GGC ACC TAC CAC GTG CGC CTG GAG GTC AAC      3454
His Thr Tyr Ala Ser Arg Gly Thr Tyr His Val Arg Leu Glu Val Asn
                1140                1145                1150

AAC ACG GTG AGC GGT GCG GCG GCC CAG GCG GAT GTG CGC GTC TTT GAG      3502
Asn Thr Val Ser Gly Ala Ala Ala Gln Ala Asp Val Arg Val Phe Glu
            1155                1160                1165

GAG CTC CGC GGA CTC AGC GTG GAC ATG AGC CTG GCC GTG GAG CAG GGC      3550
Glu Leu Arg Gly Leu Ser Val Asp Met Ser Leu Ala Val Glu Gln Gly
        1170                1175                1180

GCC CCC GTG GTG GTC AGC GCC GCG GTG CAG ACG GGC GAC AAC ATC ACG      3598
Ala Pro Val Val Val Ser Ala Ala Val Gln Thr Gly Asp Asn Ile Thr
    1185                1190                1195
```

```
TGG ACC TTC GAC ATG GGG GAC GGC ACC GTG CTG TCG GGC CCG GAG GCA      3646
Trp Thr Phe Asp Met Gly Asp Gly Thr Val Leu Ser Gly Pro Glu Ala
1200                1205                1210                1215

ACA GTG GAG CAT GTG TAC CTG CGG GCA CAG AAC TGC ACA GTG ACC GTG      3694
Thr Val Glu His Val Tyr Leu Arg Ala Gln Asn Cys Thr Val Thr Val
                    1220                1225                1230

GGT GCG GCC AGC CCC GCC GGC CAC CTG GCC CGG AGC CTG CAC GTG CTG      3742
Gly Ala Ala Ser Pro Ala Gly His Leu Ala Arg Ser Leu His Val Leu
            1235                1240                1245

GTC TTC GTC CTG GAG GTG CTG CGC GTT GAA CCC GCC GCC TGC ATC CCC      3790
Val Phe Val Leu Glu Val Leu Arg Val Glu Pro Ala Ala Cys Ile Pro
        1250                1255                1260

ACG CAG CCT GAC GCG CGG CTC ACG GCC TAC GTC ACC GGG AAC CCG GCC      3838
Thr Gln Pro Asp Ala Arg Leu Thr Ala Tyr Val Thr Gly Asn Pro Ala
    1265                1270                1275

CAC TAC CTC TTC GAC TGG ACC TTC GGG GAT GGC TCC TCC AAC ACG ACC      3886
His Tyr Leu Phe Asp Trp Thr Phe Gly Asp Gly Ser Ser Asn Thr Thr
1280                1285                1290                1295

GTG CGG GGG TGC CCG ACG GTG ACA CAC AAC TTC ACG CGG AGC GGC ACG      3934
Val Arg Gly Cys Pro Thr Val Thr His Asn Phe Thr Arg Ser Gly Thr
                    1300                1305                1310

TTC CCC CTG GCG CTG GTG CTG TCC AGC CGC GTG AAC AGG GCG CAT TAC      3982
Phe Pro Leu Ala Leu Val Leu Ser Ser Arg Val Asn Arg Ala His Tyr
            1315                1320                1325

TTC ACC AGC ATC TGC GTG GAG CCA GAG GTG GGC AAC GTC ACC CTG CAG      4030
Phe Thr Ser Ile Cys Val Glu Pro Glu Val Gly Asn Val Thr Leu Gln
        1330                1335                1340

CCA GAG AGG CAG TTT GTG CAG CTC GGG GAC GAG GCC TGG CTG GTG GCA      4078
Pro Glu Arg Gln Phe Val Gln Leu Gly Asp Glu Ala Trp Leu Val Ala
    1345                1350                1355

TGT GCC TGG CCC CCG TTC CCC TAC CGC TAC ACC TGG GAC TTT GGC ACC      4126
Cys Ala Trp Pro Pro Phe Pro Tyr Arg Tyr Thr Trp Asp Phe Gly Thr
1360                1365                1370                1375

GAG GAA GCC GCC CCC ACC CGT GCC AGG GGC CCT GAG GTG ACG TTC ATC      4174
Glu Glu Ala Ala Pro Thr Arg Ala Arg Gly Pro Glu Val Thr Phe Ile
                    1380                1385                1390

TAC CGA GAC CCA GGC TCC TAT CTT GTG ACA GTC ACC GCG TCC AAC AAC      4222
Tyr Arg Asp Pro Gly Ser Tyr Leu Val Thr Val Thr Ala Ser Asn Asn
            1395                1400                1405

ATC TCT GCT GCC AAT GAC TCA GCC CTG GTG GAG GTG CAG GAG CCC GTG      4270
Ile Ser Ala Ala Asn Asp Ser Ala Leu Val Glu Val Gln Glu Pro Val
        1410                1415                1420

CTG GTC ACC AGC ATC AAG GTC AAT GGC TCC CTT GGG CTG GAG CTG CAG      4318
Leu Val Thr Ser Ile Lys Val Asn Gly Ser Leu Gly Leu Glu Leu Gln
    1425                1430                1435
```

```
CAG CCG TAC CTG TTC TCT GCT GTG GGC CGT GGG CGC CCC GCC AGC TAC      4366
Gln Pro Tyr Leu Phe Ser Ala Val Gly Arg Gly Arg Pro Ala Ser Tyr
1440                1445                1450                1455

CTG TGG GAT CTG GGG GAC GGT GGG TGG CTC GAG GGT CCG GAG GTC ACC      4414
Leu Trp Asp Leu Gly Asp Gly Gly Trp Leu Glu Gly Pro Glu Val Thr
                    1460                1465                1470

CAC GCT TAC AAC AGC ACA GGT GAC TTC ACC GTT AGG GTG GCC GGC TGG      4462
His Ala Tyr Asn Ser Thr Gly Asp Phe Thr Val Arg Val Ala Gly Trp
            1475                1480                1485

AAT GAG GTG AGC CGC AGC GAG GCC TGG CTC AAT GTG ACG GTG AAG CGG      4510
Asn Glu Val Ser Arg Ser Glu Ala Trp Leu Asn Val Thr Val Lys Arg
        1490                1495                1500

CGC GTG CGG GGG CTC GTC GTC AAT GCA AGC CGC ACG GTG GTG CCC CTG      4558
Arg Val Arg Gly Leu Val Val Asn Ala Ser Arg Thr Val Val Pro Leu
    1505                1510                1515

AAT GGG AGC GTG AGC TTC AGC ACG TCG CTG GAG GCC GGC AGT GAT GTG      4606
Asn Gly Ser Val Ser Phe Ser Thr Ser Leu Glu Ala Gly Ser Asp Val
1520                1525                1530                1535

CGC TAT TCC TGG GTG CTC TGT GAC CGC TGC ACG CCC ATC CCT GGG GGT      4654
Arg Tyr Ser Trp Val Leu Cys Asp Arg Cys Thr Pro Ile Pro Gly Gly
                    1540                1545                1550

CCT ACC ATC TCT TAC ACC TTC CGC TCC GTG GGC ACC TTC AAT ATC ATC      4702
Pro Thr Ile Ser Tyr Thr Phe Arg Ser Val Gly Thr Phe Asn Ile Ile
            1555                1560                1565

GTC ACG GCT GAG AAC GAG GTG GGC TCC GCC CAG GAC AGC ATC TTC GTC      4750
Val Thr Ala Glu Asn Glu Val Gly Ser Ala Gln Asp Ser Ile Phe Val
        1570                1575                1580

TAT GTC CTG CAG CTC ATA GAG GGG CTG CAG GTG GTG GGC GGT GGC CGC      4798
Tyr Val Leu Gln Leu Ile Glu Gly Leu Gln Val Val Gly Gly Gly Arg
    1585                1590                1595

TAC TTC CCC ACC AAC CAC ACG GTA CAG CTG CAG GCC GTG GTT AGG GAT      4846
Tyr Phe Pro Thr Asn His Thr Val Gln Leu Gln Ala Val Val Arg Asp
1600                1605                1610                1615

GGC ACC AAC GTC TCC TAC AGC TGG ACT GCC TGG AGG GAC AGG GGC CCG      4894
Gly Thr Asn Val Ser Tyr Ser Trp Thr Ala Trp Arg Asp Arg Gly Pro
                    1620                1625                1630

GCC CTG GCC GGC AGC GGC AAA GGC TTC TCG CTC ACC GTG CTC GAG GCC      4942
Ala Leu Ala Gly Ser Gly Lys Gly Phe Ser Leu Thr Val Leu Glu Ala
            1635                1640                1645

GGC ACC TAC CAT GTG CAG CTG CGG GCC ACC AAC ATG CTG GGC AGC GCC      4990
Gly Thr Tyr His Val Gln Leu Arg Ala Thr Asn Met Leu Gly Ser Ala
        1650                1655                1660

TGG GCC GAC TGC ACC ATG GAC TTC GTG GAG CCT GTG GGG TGG CTG ATG      5038
Trp Ala Asp Cys Thr Met Asp Phe Val Glu Pro Val Gly Trp Leu Met
    1665                1670                1675
```

```
GTG ACC GCC TCC CCG AAC CCA GCT GCC GTC AAC ACA AGC GTC ACC CTC      5086
Val Thr Ala Ser Pro Asn Pro Ala Ala Val Asn Thr Ser Val Thr Leu
1680                1685                1690                1695

AGT GCC GAG CTG GCT GGT GGC AGT GGT GTC GTA TAC ACT TGG TCC TTG      5134
Ser Ala Glu Leu Ala Gly Gly Ser Gly Val Val Tyr Thr Trp Ser Leu
                    1700                1705                1710

GAG GAG GGG CTG AGC TGG GAG ACC TCC GAG CCA TTT ACC ACC CAT AGC      5182
Glu Glu Gly Leu Ser Trp Glu Thr Ser Glu Pro Phe Thr Thr His Ser
            1715                1720                1725

TTC CCC ACA CCC GGC CTG CAC TTG GTC ACC ATG ACG GCA GGG AAC CCG      5230
Phe Pro Thr Pro Gly Leu His Leu Val Thr Met Thr Ala Gly Asn Pro
        1730                1735                1740

CTG GGC TCA GCC AAC GCC ACC GTG GAA GTG GAT GTG CAG GTG CCT GTG      5278
Leu Gly Ser Ala Asn Ala Thr Val Glu Val Asp Val Gln Val Pro Val
    1745                1750                1755

AGT GGC CTC AGC ATC AGG GCC AGC GAG CCC GGA GGC AGC TTC GTG GCG      5326
Ser Gly Leu Ser Ile Arg Ala Ser Glu Pro Gly Gly Ser Phe Val Ala
1760                1765                1770                1775

GCC GGG TCC TCT GTG CCC TTT TGG GGG CAG CTG GCC ACG GGC ACC AAT      5374
Ala Gly Ser Ser Val Pro Phe Trp Gly Gln Leu Ala Thr Gly Thr Asn
                    1780                1785                1790

GTG AGC TGG TGC TGG GCT GTG CCC GGC GGC AGC AGC AAG CGT GGC CCT      5422
Val Ser Trp Cys Trp Ala Val Pro Gly Gly Ser Ser Lys Arg Gly Pro
            1795                1800                1805

CAT GTC ACC ATG GTC TTC CCG GAT GCT GGC ACC TTC TCC ATC CGG CTC      5470
His Val Thr Met Val Phe Pro Asp Ala Gly Thr Phe Ser Ile Arg Leu
        1810                1815                1820

AAT GCC TCC AAC GCA GTC AGC TGG GTC TCA GCC ACG TAC AAC CTC ACG      5518
Asn Ala Ser Asn Ala Val Ser Trp Val Ser Ala Thr Tyr Asn Leu Thr
    1825                1830                1835

GCG GAG GAG CCC ATC GTG GGC CTG GTG CTG TGG GCC AGC AGC AAG GTG      5566
Ala Glu Glu Pro Ile Val Gly Leu Val Leu Trp Ala Ser Ser Lys Val
1840                1845                1850                1855

GTG GCG CCC GGG CAG CTG GTC CAT TTT CAG ATC CTG CTG GCT GCC GGC      5614
Val Ala Pro Gly Gln Leu Val His Phe Gln Ile Leu Leu Ala Ala Gly
                    1860                1865                1870

TCA GCT GTC ACC TTC CGC CTG CAG GTC GGC GGG GCC AAC CCC GAG GTG      5662
Ser Ala Val Thr Phe Arg Leu Gln Val Gly Gly Ala Asn Pro Glu Val
            1875                1880                1885

CTC CCC GGG CCC CGT TTC TCC CAC AGC TTC CCC CGC GTC GGA GAC CAC      5710
Leu Pro Gly Pro Arg Phe Ser His Ser Phe Pro Arg Val Gly Asp His
        1890                1895                1900

GTG GTG AGC GTG CGG GGC AAA AAC CAC GTG AGC TGG GCC CAG GCG CAG      5758
Val Val Ser Val Arg Gly Lys Asn His Val Ser Trp Ala Gln Ala Gln
    1905                1910                1915
```

```
GTG CGC ATC GTG GTG CTG GAG GCC GTG AGT GGG CTG CAG ATG CCC AAC      5806
Val Arg Ile Val Val Leu Glu Ala Val Ser Gly Leu Gln Met Pro Asn
1920                1925                1930                1935

TGC TGC GAG CCT GGC ATC GCC ACG GGC ACT GAG AGG AAC TTC ACA GCC      5854
Cys Cys Glu Pro Gly Ile Ala Thr Gly Thr Glu Arg Asn Phe Thr Ala
                    1940                1945                1950

CGC GTG CAG CGC GGC TCT CGG GTC GCC TAC GCC TGG TAC TTC TCG CTG      5902
Arg Val Gln Arg Gly Ser Arg Val Ala Tyr Ala Trp Tyr Phe Ser Leu
                1955                1960                1965

CAG AAG GTC CAG GGC GAC TCG CTG GTC ATC CTG TCG GGC CGC GAC GTC      5950
Gln Lys Val Gln Gly Asp Ser Leu Val Ile Leu Ser Gly Arg Asp Val
        1970                1975                1980

ACC TAC ACG CCC GTG GCC GCG GGG CTG TTG GAG ATC CAG GTG CGC GCC      5998
Thr Tyr Thr Pro Val Ala Ala Gly Leu Leu Glu Ile Gln Val Arg Ala
    1985                1990                1995

TTC AAC GCC CTG GGC AGT GAG AAC CGC ACG CTG GTG CTG GAG GTT CAG      6046
Phe Asn Ala Leu Gly Ser Glu Asn Arg Thr Leu Val Leu Glu Val Gln
2000                2005                2010                2015

GAC GCC GTC CAG TAT GTG GCC CTG CAG AGC GGC CCC TGC TTC ACC AAC      6094
Asp Ala Val Gln Tyr Val Ala Leu Gln Ser Gly Pro Cys Phe Thr Asn
                    2020                2025                2030

CGC TCG GCG CAG TTT GAG GCC GCC ACC AGC CGC AGC CCC CGG CGT GTG      6142
Arg Ser Ala Gln Phe Glu Ala Ala Thr Ser Pro Ser Pro Arg Arg Val
                2035                2040                2045

GCC TAC CAC TGG GAC TTT GGG GAT GGG TCG CCA GGG CAG GAC ACA GAT      6190
Ala Tyr His Trp Asp Phe Gly Asp Gly Ser Pro Gly Gln Asp Thr Asp
        2050                2055                2060

GAG CCC AGG GCC GAG CAC TCC TAC CTG AGG CCT GGG GAC TAC CGC GTG      6238
Glu Pro Arg Ala Glu His Ser Tyr Leu Arg Pro Gly Asp Tyr Arg Val
    2065                2070                2075

CAG GTG AAC GCC TCC AAC CTG GTG AGC TTC TTC GTG GCG CAG GCC ACG      6286
Gln Val Asn Ala Ser Asn Leu Val Ser Phe Phe Val Ala Gln Ala Thr
2080                2085                2090                2095

GTG ACC GTC CAG GTG CTG GCC TGC CGG GAG CCG GAG GTG GAC GTG GTC      6334
Val Thr Val Gln Val Leu Ala Cys Arg Glu Pro Glu Val Asp Val Val
                    2100                2105                2110

CTG CCC CTG CAG GTG CTG ATG CGG CGA TCA CAG CGC AAC TAC TTG GAG      6382
Leu Pro Leu Gln Val Leu Met Arg Arg Ser Gln Arg Asn Tyr Leu Glu
                2115                2120                2125

GCC CAC GTT GAC CTG CGC GAC TGC GTC ACC TAC CAG ACT GAG TAC CGC      6430
Ala His Val Asp Leu Arg Asp Cys Val Thr Tyr Gln Thr Glu Tyr Arg
        2130                2135                2140

TGG GAG GTG TAT CGC ACC GCC AGC TGC CAG CGG CCG GGG CGC CCA GCG      6478
Trp Glu Val Tyr Arg Thr Ala Ser Cys Gln Arg Pro Gly Arg Pro Ala
    2145                2150                2155
```

```
CGT GTG GCC CTG CCC GGC GTG GAC GTG AGC CGG CCT CGG CTG GTG CTG      6526
Arg Val Ala Leu Pro Gly Val Asp Val Ser Arg Pro Arg Leu Val Leu
2160                2165                2170                2175

CCG CGG CTG GCG CTG CCT GTG GGG CAC TAC TGC TTT GTG TTT GTC GTG      6574
Pro Arg Leu Ala Leu Pro Val Gly His Tyr Cys Phe Val Phe Val Val
                    2180                2185                2190

TCA TTT GGG GAC ACG CCA CTG ACA CAG AGC ATC CAG GCC AAT GTG ACG      6622
Ser Phe Gly Asp Thr Pro Leu Thr Gln Ser Ile Gln Ala Asn Val Thr
                2195                2200                2205

GTG GCC CCC GAG CGC CTG GTG CCC ATC ATT GAG GGT GGC TCA TAC CGC      6670
Val Ala Pro Glu Arg Leu Val Pro Ile Ile Glu Gly Gly Ser Tyr Arg
        2210                2215                2220

GTG TGG TCA GAC ACA CGG GAC CTG GTG CTG GAT GGG AGC GAG TCC TAC      6718
Val Trp Ser Asp Thr Arg Asp Leu Val Leu Asp Gly Ser Glu Ser Tyr
    2225                2230                2235

GAC CCC AAC CTG GAG GAC GGC GAC CAG ACG CCG CTC AGT TTC CAC TGG      6766
Asp Pro Asn Leu Glu Asp Gly Asp Gln Thr Pro Leu Ser Phe His Trp
2240                2245                2250                2255

GCC TGT GTG GCT TCG ACA CAG AGG GAG GCT GGC GGG TGT GCG CTG AAC      6814
Ala Cys Val Ala Ser Thr Gln Arg Glu Ala Gly Gly Cys Ala Leu Asn
                2260                2265                2270

TTT GGG CCC CGC GGG AGC AGC ACG GTC ACC ATT CCA CGG GAG CGG CTG      6862
Phe Gly Pro Arg Gly Ser Ser Thr Val Thr Ile Pro Arg Glu Arg Leu
            2275                2280                2285

GCG GCT GGC GTG GAG TAC ACC TTC AGC CTG ACC GTG TGG AAG GCC GGC      6910
Ala Ala Gly Val Glu Tyr Thr Phe Ser Leu Thr Val Trp Lys Ala Gly
        2290                2295                2300

CGC AAG GAG GAG GCC ACC AAC CAG ACG GTG CTG ATC CGG AGT GGC CGG      6958
Arg Lys Glu Glu Ala Thr Asn Gln Thr Val Leu Ile Arg Ser Gly Arg
    2305                2310                2315

GTG CCC ATT GTG TCC TTG GAG TGT GTG TCC TGC AAG GCA CAG GCC GTG      7006
Val Pro Ile Val Ser Leu Glu Cys Val Ser Cys Lys Ala Gln Ala Val
2320                2325                2330                2335

TAC GAA GTG AGC CGC AGC TCC TAC GTG TAC TTG GAG GGC CGC TGC CTC      7054
Tyr Glu Val Ser Arg Ser Ser Tyr Val Tyr Leu Glu Gly Arg Cys Leu
                2340                2345                2350

AAT TGC AGC AGC GGC TCC AAG CGA GGG CGG TGG GCT GCA CGT ACG TTC      7102
Asn Cys Ser Ser Gly Ser Lys Arg Gly Arg Trp Ala Ala Arg Thr Phe
            2355                2360                2365

AGC AAC AAG ACG CTG GTG CTG GAT GAG ACC ACC ACA TCC ACG GGC AGT      7150
Ser Asn Lys Thr Leu Val Leu Asp Glu Thr Thr Thr Ser Thr Gly Ser
        2370                2375                2380

GCA GGC ATG CGA CTG GTG CTG CGG CGG GGC GTG CTG CGG GAC GGC GAG      7198
Ala Gly Met Arg Leu Val Leu Arg Arg Gly Val Leu Arg Asp Gly Glu
    2385                2390                2395
```

```
GGA TAC ACC TTC ACG CTC ACG GTG CTG GGC CGC TCT GGC GAG GAG GAG      7246
Gly Tyr Thr Phe Thr Leu Thr Val Leu Gly Arg Ser Gly Glu Glu Glu
2400                2405                2410                2415

GGC TGC GCC TCC ATC CGC CTG TCC CCC AAC CGC CCG CCG CTG GGG GGC      7294
Gly Cys Ala Ser Ile Arg Leu Ser Pro Asn Arg Pro Pro Leu Gly Gly
                    2420                2425                2430

TCT TGC CGC CTC TTC CCA CTG GGC GCT GTG CAC GCC CTC ACC ACC AAG      7342
Ser Cys Arg Leu Phe Pro Leu Gly Ala Val His Ala Leu Thr Thr Lys
                2435                2440                2445

GTG CAC TTC GAA TGC ACG GGC TGG CAT GAC GCG GAG GAT GCT GGC GCC      7390
Val His Phe Glu Cys Thr Gly Trp His Asp Ala Glu Asp Ala Gly Ala
        2450                2455                2460

CCG CTG GTG TAC GCC CTG CTG CTG CGG CGC TGT CGC CAG GGC CAC TGC      7438
Pro Leu Val Tyr Ala Leu Leu Leu Arg Arg Cys Arg Gln Gly His Cys
    2465                2470                2475

GAG GAG TTC TGT GTC TAC AAG GGC AGC CTC TCC AGC TAC GGA GCC GTG      7486
Glu Glu Phe Cys Val Tyr Lys Gly Ser Leu Ser Ser Tyr Gly Ala Val
2480                2485                2490                2495

CTG CCC CCG GGT TTC AGG CCA CAC TTC GAG GTG GGC CTG GCC GTG GTG      7534
Leu Pro Pro Gly Phe Arg Pro His Phe Glu Val Gly Leu Ala Val Val
                2500                2505                2510

GTG CAG GAC CAG CTG GGA GCC GCT GTG GTC GCC CTC AAC AGG TCT TTG      7582
Val Gln Asp Gln Leu Gly Ala Ala Val Val Ala Leu Asn Arg Ser Leu
            2515                2520                2525

GCC ATC ACC CTC CCA GAG CCC AAC GGC AGC GCA ACG GGG CTC ACA GTC      7630
Ala Ile Thr Leu Pro Glu Pro Asn Gly Ser Ala Thr Gly Leu Thr Val
        2530                2535                2540

TGG CTG CAC GGG CTC ACC GCT AGT GTG CTC CCA GGG CTG CTG CGG CAG      7678
Trp Leu His Gly Leu Thr Ala Ser Val Leu Pro Gly Leu Leu Arg Gln
    2545                2550                2555

GCC GAT CCC CAG CAC GTC ATC GAG TAC TCG TTG GCC CTG GTC ACC GTG      7726
Ala Asp Pro Gln His Val Ile Glu Tyr Ser Leu Ala Leu Val Thr Val
2560                2565                2570                2575

CTG AAC GAG TAC GAG CGG GCC CTG GAC GTG GCG GCA GAG CCC AAG CAC      7774
Leu Asn Glu Tyr Glu Arg Ala Leu Asp Val Ala Ala Glu Pro Lys His
                2580                2585                2590

GAG CGG CAG CAC CGA GCC CAG ATA CGC AAG AAC ATC ACG GAG ACT CTG      7822
Glu Arg Gln His Arg Ala Gln Ile Arg Lys Asn Ile Thr Glu Thr Leu
            2595                2600                2605

GTG TCC CTG AGG GTC CAC ACT GTG GAT GAC ATC CAG CAG ATC GCT GCT      7870
Val Ser Leu Arg Val His Thr Val Asp Asp Ile Gln Gln Ile Ala Ala
        2610                2615                2620

GCG CTG GCC CAG TGC ATG GGG CCC AGC AGG GAG CTC GTA TGC CGC TCG      7918
Ala Leu Ala Gln Cys Met Gly Pro Ser Arg Glu Leu Val Cys Arg Ser
    2625                2630                2635
```

```
TGC CTG AAG CAG ACG CTG CAC AAG CTG GAG GCC ATG ATG CTC ATC CTG      7966
Cys Leu Lys Gln Thr Leu His Lys Leu Glu Ala Met Met Leu Ile Leu
2640                2645                2650                2655

CAG GCA GAG ACC ACC GCG GGC ACC GTG ACG CCC ACC GCC ATC GGA GAC      8014
Gln Ala Glu Thr Thr Ala Gly Thr Val Thr Pro Thr Ala Ile Gly Asp
                    2660                2665                2670

AGC ATC CTC AAC ATC ACA GGA GAC CTC ATC CAC CTG GCC AGC TCG GAC      8062
Ser Ile Leu Asn Ile Thr Gly Asp Leu Ile His Leu Ala Ser Ser Asp
                2675                2680                2685

GTG CGG GCA CCA CAG CCC TCA GAG CTG GGA GCC GAG TCA CCA TCT CGG      8110
Val Arg Ala Pro Gln Pro Ser Glu Leu Gly Ala Glu Ser Pro Ser Arg
        2690                2695                2700

ATG GTG GCG TCC CAG GCC TAC AAC CTG ACC TCT GCC CTC ATG CGC ATC      8158
Met Val Ala Ser Gln Ala Tyr Asn Leu Thr Ser Ala Leu Met Arg Ile
    2705                2710                2715

CTC ATG CGC TCC CGC GTG CTC AAC GAG GAG CCC CTG ACG CTG GCG GGC      8206
Leu Met Arg Ser Arg Val Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly
2720                2725                2730                2735

GAG GAG ATC GTG GCC CAG GGC AAG CGC TCG GAC CCG CGG AGC CTG CTG      8254
Glu Glu Ile Val Ala Gln Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu
                    2740                2745                2750

TGC TAT GGC GGC GCC CCA GGG CCT GGC TGC CAC TTC TCC ATC CCC GAG      8302
Cys Tyr Gly Gly Ala Pro Gly Pro Gly Cys His Phe Ser Ile Pro Glu
                2755                2760                2765

GCT TTC AGC GGG GCC CTG GCC AAC CTC AGT GAC GTG GTG CAG CTC ATC      8350
Ala Phe Ser Gly Ala Leu Ala Asn Leu Ser Asp Val Val Gln Leu Ile
        2770                2775                2780

TTT CTG GTG GAC TCC AAT CCC TTT CCC TTT GGC TAT ATC AGC AAC TAC      8398
Phe Leu Val Asp Ser Asn Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr
    2785                2790                2795

ACC GTC TCC ACC AAG GTG GCC TCG ATG GCA TTC CAG ACA CAG GCC GGC      8446
Thr Val Ser Thr Lys Val Ala Ser Met Ala Phe Gln Thr Gln Ala Gly
2800                2805                2810                2815

GCC CAG ATC CCC ATC GAG CGG CTG GCC TCA GAG CGC GCC ATC ACC GTG      8494
Ala Gln Ile Pro Ile Glu Arg Leu Ala Ser Glu Arg Ala Ile Thr Val
                    2820                2825                2830

AAG GTG CCC AAC AAC TCG GAC TGG GCT GCC CGG GGC CAC CGC AGC TCC      8542
Lys Val Pro Asn Asn Ser Asp Trp Ala Ala Arg Gly His Arg Ser Ser
                2835                2840                2845

GCC AAC TCC GCC AAC TCC GTT GTG GTC CAG CCC CAG GCC TCC GTC GGT      8590
Ala Asn Ser Ala Asn Ser Val Val Val Gln Pro Gln Ala Ser Val Gly
        2850                2855                2860

GCT GTG GTC ACC CTG GAC AGC AGC AAC CCT GCG GCC GGG CTG CAT CTG      8638
Ala Val Val Thr Leu Asp Ser Ser Asn Pro Ala Ala Gly Leu His Leu
    2865                2870                2875
```

```
CAG CTC AAC TAT ACG CTG CTG GAC GGC CAC TAC CTG TCT GAG GAA CCT      8686
Gln Leu Asn Tyr Thr Leu Leu Asp Gly His Tyr Leu Ser Glu Glu Pro
2880                2885                2890                2895

GAG CCC TAC CTG GCA GTC TAC CTA CAC TCG GAG CCC CGG CCC AAT GAG      8734
Glu Pro Tyr Leu Ala Val Tyr Leu His Ser Glu Pro Arg Pro Asn Glu
                    2900                2905                2910

CAC AAC TGC TCG GCT AGC AGG AGG ATC CGC CCA GAG TCA CTC CAG GGT      8782
His Asn Cys Ser Ala Ser Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly
            2915                2920                2925

GCT GAC CAC CGG CCC TAC ACC TTC TTC ATT TCC CCG GGG AGC AGA GAC      8830
Ala Asp His Arg Pro Tyr Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp
        2930                2935                2940

CCA GCG GGG AGT TAC CAT CTG AAC CTC TCC AGC CAC TTC CGC TGG TCG      8878
Pro Ala Gly Ser Tyr His Leu Asn Leu Ser Ser His Phe Arg Trp Ser
    2945                2950                2955

GCG CTG CAG GTG TCC GTG GGC CTG TAC ACG TCC CTG TGC CAG TAC TTC      8926
Ala Leu Gln Val Ser Val Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe
2960                2965                2970                2975

AGC GAG GAG GAC ATG GTG TGG CGG ACA GAG GGG CTG CTG CCC CTG GAG      8974
Ser Glu Glu Asp Met Val Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu
                    2980                2985                2990

GAG ACC TCG CCC CGC CAG GCC GTC TGC CTC ACC CGC CAC CTC ACC GCC      9022
Glu Thr Ser Pro Arg Gln Ala Val Cys Leu Thr Arg His Leu Thr Ala
            2995                3000                3005

TTC GGC GCC AGC CTC TTC GTG CCC CCA AGC CAT GTC CGC TTT GTG TTT      9070
Phe Gly Ala Ser Leu Phe Val Pro Pro Ser His Val Arg Phe Val Phe
        3010                3015                3020

CCT GAG CCG ACA GCG GAT GTA AAC TAC ATC GTC ATG CTG ACA TGT GCT      9118
Pro Glu Pro Thr Ala Asp Val Asn Tyr Ile Val Met Leu Thr Cys Ala
    3025                3030                3035

GTG TGC CTG GTG ACC TAC ATG GTC ATG GCC GCC ATC CTG CAC AAG CTG      9166
Val Cys Leu Val Thr Tyr Met Val Met Ala Ala Ile Leu His Lys Leu
3040                3045                3050                3055

GAC CAG TTG GAT GCC AGC CGG GGC CGC GCC ATC CCT TTC TGT GGG CAG      9214
Asp Gln Leu Asp Ala Ser Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln
                    3060                3065                3070

CGG GGC CGC TTC AAG TAC GAG ATC CTC GTC AAG ACA GGC TGG GGC CGG      9262
Arg Gly Arg Phe Lys Tyr Glu Ile Leu Val Lys Thr Gly Trp Gly Arg
            3075                3080                3085

GGC TCA GGT ACC ACG GCC CAC GTG GGC ATC ATG CTG TAT GGG GTG GAC      9310
Gly Ser Gly Thr Thr Ala His Val Gly Ile Met Leu Tyr Gly Val Asp
        3090                3095                3100

AGC CGG AGC GGC CAC CGG CAC CTG GAC GGC GAC AGA GCC TTC CAC CGC      9358
Ser Arg Ser Gly His Arg His Leu Asp Gly Asp Arg Ala Phe His Arg
    3105                3110                3115
```

```
AAC AGC CTG GAC ATC TTC CGG ATC GCC ACC CCG CAC AGC CTG GGT AGC      9406
Asn Ser Leu Asp Ile Phe Arg Ile Ala Thr Pro His Ser Leu Gly Ser
3120                 3125                 3130                 3135

GTG TGG AAG ATC CGA GTG TGG CAC GAC AAC AAA GGG CTC AGC CCT GCC      9454
Val Trp Lys Ile Arg Val Trp His Asp Asn Lys Gly Leu Ser Pro Ala
                3140                 3145                 3150

TGG TTC CTG CAG CAC GTC ATC GTC AGG GAC CTG CAG ACG GCA CGC AGC      9502
Trp Phe Leu Gln His Val Ile Val Arg Asp Leu Gln Thr Ala Arg Ser
           3155                 3160                 3165

GCC TTC TTC CTG GTC AAT GAC TGG CTT TCG GTG GAG ACG GAG GCC AAC      9550
Ala Phe Phe Leu Val Asn Asp Trp Leu Ser Val Glu Thr Glu Ala Asn
      3170                 3175                 3180

GGG GGC CTG GTG GAG AAG GAG GTG CTG GCC GCG AGC GAC GCA GCC CTT      9598
Gly Gly Leu Val Glu Lys Glu Val Leu Ala Ala Ser Asp Ala Ala Leu
 3185                 3190                 3195

TTG CGC TTC CGG CGC CTG CTG GTG GCT GAG CTG CAG CGT GGC TTC TTT      9646
Leu Arg Phe Arg Arg Leu Leu Val Ala Glu Leu Gln Arg Gly Phe Phe
3200                 3205                 3210                 3215

GAC AAG CAC ATC TGG CTC TCC ATA TGG GAC CGG CCG CCT CGT AGC CGT      9694
Asp Lys His Ile Trp Leu Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg
                3220                 3225                 3230

TTC ACT CGC ATC CAG AGG GCC ACC TGC TGC GTT CTC CTC ATC TGC CTC      9742
Phe Thr Arg Ile Gln Arg Ala Thr Cys Cys Val Leu Leu Ile Cys Leu
           3235                 3240                 3245

TTC CTG GGC GCC AAC GCC GTG TGG TAC GGG GCT GTT GGC GAC TCT GCC      9790
Phe Leu Gly Ala Asn Ala Val Trp Tyr Gly Ala Val Gly Asp Ser Ala
      3250                 3255                 3260

TAC AGC ACG GGG CAT GTG TCC AGG CTG AGC CCG CTG AGC GTC GAC ACA      9838
Tyr Ser Thr Gly His Val Ser Arg Leu Ser Pro Leu Ser Val Asp Thr
 3265                 3270                 3275

GTC GCT GTT GGC CTG GTG TCC AGC GTG GTT GTC TAT CCC GTC TAC CTG      9886
Val Ala Val Gly Leu Val Ser Ser Val Val Val Tyr Pro Val Tyr Leu
3280                 3285                 3290                 3295

GCC ATC CTT TTT CTC TTC CGG ATG TCC CGG AGC AAG GTG GCT GGG AGC      9934
Ala Ile Leu Phe Leu Phe Arg Met Ser Arg Ser Lys Val Ala Gly Ser
                3300                 3305                 3310

CCG AGC CCC ACA CCT GCC GGG CAG CAG GTG CTG GAC ATC GAC AGC TGC      9982
Pro Ser Pro Thr Pro Ala Gly Gln Gln Val Leu Asp Ile Asp Ser Cys
           3315                 3320                 3325

CTG GAC TCG TCC GTG CTG GAC AGC TCC TTC CTC ACG TTC TCA GGC CTC     10030
Leu Asp Ser Ser Val Leu Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu
      3330                 3335                 3340

CAC GCT GAG GCC TTT GTT GGA CAG ATG AAG AGT GAC TTG TTT CTG GAT     10078
His Ala Glu Ala Phe Val Gly Gln Met Lys Ser Asp Leu Phe Leu Asp
 3345                 3350                 3355
```

```
GAT TCT AAG AGT CTG GTG TGC TGG CCC TCC GGC GAG GGA ACG CTC AGT      10126
Asp Ser Lys Ser Leu Val Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser
3360                 3365                 3370                 3375

TGG CCG GAC CTG CTC AGT GAC CCG TCC ATT GTG GGT AGC AAT CTG CGG      10174
Trp Pro Asp Leu Leu Ser Asp Pro Ser Ile Val Gly Ser Asn Leu Arg
                3380                 3385                 3390

CAG CTG GCA CGG GGC CAG GCG GGC CAT GGG CTG GGC CCA GAG GAG GAC      10222
Gln Leu Ala Arg Gly Gln Ala Gly His Gly Leu Gly Pro Glu Glu Asp
            3395                 3400                 3405

GGC TTC TCC CTG GCC AGC CCC TAC TCG CCT GCC AAA TCC TTC TCA GCA      10270
Gly Phe Ser Leu Ala Ser Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala
        3410                 3415                 3420

TCA GAT GAA GAC CTG ATC CAG CAG GTC CTT GCC GAG GGG GTC AGC AGC      10318
Ser Asp Glu Asp Leu Ile Gln Gln Val Leu Ala Glu Gly Val Ser Ser
    3425                 3430                 3435

CCA GCC CCT ACC CAA GAC ACC CAC ATG GAA ACG GAC CTG CTC AGC AGC      10366
Pro Ala Pro Thr Gln Asp Thr His Met Glu Thr Asp Leu Leu Ser Ser
3440                 3445                 3450                 3455

CTG TCC AGC ACT CCT GGG GAG AAG ACA GAG ACG CTG GCG CTG CAG AGG      10414
Leu Ser Ser Thr Pro Gly Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg
                3460                 3465                 3470

CTG GGG GAG CTG GGG CCA CCC AGC CCA GGC CTG AAC TGG GAA CAG CCC      10462
Leu Gly Glu Leu Gly Pro Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro
            3475                 3480                 3485

CAG GCA GCG AGG CTG TCC AGG ACA GGA CTG GTG GAG GGT CTG CGG AAG      10510
Gln Ala Ala Arg Leu Ser Arg Thr Gly Leu Val Glu Gly Leu Arg Lys
        3490                 3495                 3500

CGC CTG CTG CCG GCC TGG TGT GCC TCC CTG GCC CAC GGG CTC AGC CTG      10558
Arg Leu Leu Pro Ala Trp Cys Ala Ser Leu Ala His Gly Leu Ser Leu
    3505                 3510                 3515

CTC CTG GTG GCT GTG GCT GTG GCT GTC TCA GGG TGG GTG GGT GCG AGC      10606
Leu Leu Val Ala Val Ala Val Ala Val Ser Gly Trp Val Gly Ala Ser
3520                 3525                 3530                 3535

TTC CCC CCG GGC GTG AGT GTT GCG TGG CTC CTG TCC AGC AGC GCC AGC      10654
Phe Pro Pro Gly Val Ser Val Ala Trp Leu Leu Ser Ser Ser Ala Ser
                3540                 3545                 3550

TTC CTG GCC TCA TTC CTC GGC TGG GAG CCA CTG AAG GTC TTG CTG GAA      10702
Phe Leu Ala Ser Phe Leu Gly Trp Glu Pro Leu Lys Val Leu Leu Glu
            3555                 3560                 3565

GCC CTG TAC TTC TCA CTG GTG GCC AAG CGG CTG CAC CCG GAT GAA GAT      10750
Ala Leu Tyr Phe Ser Leu Val Ala Lys Arg Leu His Pro Asp Glu Asp
        3570                 3575                 3580

GAC ACC CTG GTA GAG AGC CCG GCT GTG ACG CCT GTG AGC GCA CGT GTG      10798
Asp Thr Leu Val Glu Ser Pro Ala Val Thr Pro Val Ser Ala Arg Val
    3585                 3590                 3595
```

```
CCC CGC GTA CGG CCA CCC CAC GGC TTT GCA CTC TTC CTG GCC AAG GAA      10846
Pro Arg Val Arg Pro Pro His Gly Phe Ala Leu Phe Leu Ala Lys Glu
3600                 3605                 3610                 3615

GAA GCC CGC AAG GTC AAG AGG CTA CAT GGC ATG CTG CGG AGC CTC CTG      10894
Glu Ala Arg Lys Val Lys Arg Leu His Gly Met Leu Arg Ser Leu Leu
                3620                 3625                 3630

GTG TAC ATG CTT TTT CTG CTG GTG ACC CTG CTG GCC AGC TAT GGG GAT      10942
Val Tyr Met Leu Phe Leu Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp
            3635                 3640                 3645

GCC TCA TGC CAT GGG CAC GCC TAC CGT CTG CAA AGC GCC ATC AAG CAG      10990
Ala Ser Cys His Gly His Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln
        3650                 3655                 3660

GAG CTG CAC AGC CGG GCC TTC CTG GCC ATC ACG CGG TCT GAG GAG CTC      11038
Glu Leu His Ser Arg Ala Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu
    3665                 3670                 3675

TGG CCA TGG ATG GCC CAC GTG CTG CTG CCC TAC GTC CAC GGG AAC CAG      11086
Trp Pro Trp Met Ala His Val Leu Leu Pro Tyr Val His Gly Asn Gln
3680                 3685                 3690                 3695

TCC AGC CCA GAG CTG GGG CCC CCA CGG CTG CGG CAG GTG CGG CTG CAG      11134
Ser Ser Pro Glu Leu Gly Pro Pro Arg Leu Arg Gln Val Arg Leu Gln
                3700                 3705                 3710

GAA GCA CTC TAC CCA GAC CCT CCC GGC CCC AGG GTC CAC ACG TGC TCG      11182
Glu Ala Leu Tyr Pro Asp Pro Pro Gly Pro Arg Val His Thr Cys Ser
            3715                 3720                 3725

GCC GCA GGA GGC TTC AGC ACC AGC GAT TAC GAC GTT GGC TGG GAG AGT      11230
Ala Ala Gly Gly Phe Ser Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser
        3730                 3735                 3740

CCT CAC AAT GGC TCG GGG ACG TGG GCC TAT TCA GCG CCG GAT CTG CTG      11278
Pro His Asn Gly Ser Gly Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu
    3745                 3750                 3755

GGG GCA TGG TCC TGG GGC TCC TGT GCC GTG TAT GAC AGC GGG GGC TAC      11326
Gly Ala Trp Ser Trp Gly Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr
3760                 3765                 3770                 3775

GTG CAG GAG CTG GGC CTG AGC CTG GAG GAG AGC CGC GAC CGG CTG CGC      11374
Val Gln Glu Leu Gly Leu Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg
                3780                 3785                 3790

TTC CTG CAG CTG CAC AAC TGG CTG GAC AAC AGG AGC CGC GCT GTG TTC      11422
Phe Leu Gln Leu His Asn Trp Leu Asp Asn Arg Ser Arg Ala Val Phe
            3795                 3800                 3805

CTG GAG CTC ACG CGC TAC AGC CCG GCC GTG GGG CTG CAC GCC GCC GTC      11470
Leu Glu Leu Thr Arg Tyr Ser Pro Ala Val Gly Leu His Ala Ala Val
        3810                 3815                 3820

ACG CTG CGC CTC GAG TTC CCG GCG GCC GGC CGC GCC CTG GCC GCC CTC      11518
Thr Leu Arg Leu Glu Phe Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu
    3825                 3830                 3835
```

```
AGC GTC CGC CCC TTT GCG CTG CGC CGC CTC AGC GCG GGC CTC TCG CTG      11566
Ser Val Arg Pro Phe Ala Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu
3840                3845                3850                3855

CCT CTG CTC ACC TCG GTG TGC CTG CTG CTG TTC GCC GTG CAC TTC GCC      11614
Pro Leu Leu Thr Ser Val Cys Leu Leu Leu Phe Ala Val His Phe Ala
                    3860                3865                3870

GTG GCC GAG GCC CGT ACT TGG CAC AGG GAA GGG CGC TGG CGC GTG CTG      11662
Val Ala Glu Ala Arg Thr Trp His Arg Glu Gly Arg Trp Arg Val Leu
            3875                3880                3885

CGG CTC GGA GCC TGG GCG CGG TGG CTG CTG GTG GCG CTG ACG GCG GCC      11710
Arg Leu Gly Ala Trp Ala Arg Trp Leu Leu Val Ala Leu Thr Ala Ala
        3890                3895                3900

ACG GCA CTG GTA CGC CTC GCC CAG CTG GGT GCC GCT GAC CGC CAG TGG      11758
Thr Ala Leu Val Arg Leu Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp
    3905                3910                3915

ACC CGT TTC GTG CGC GGC CGC CCG CGC CGC TTC ACT AGC TTC GAC CAG      11806
Thr Arg Phe Val Arg Gly Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln
3920                3925                3930                3935

GTG GCG CAC GTG AGC TCC GCA GCC CGT GGC CTG GCG GCC TCG CTG CTC      11854
Val Ala His Val Ser Ser Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu
                    3940                3945                3950

TTC CTG CTT TTG GTC AAG GCT GCC CAG CAC GTA CGC TTC GTG CGC CAG      11902
Phe Leu Leu Leu Val Lys Ala Ala Gln His Val Arg Phe Val Arg Gln
            3955                3960                3965

TGG TCC GTC TTT GGC AAG ACA TTA TGC CGA GCT CTG CCA GAG CTC CTG      11950
Trp Ser Val Phe Gly Lys Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu
        3970                3975                3980

GGG GTC ACC TTG GGC CTG GTG GTG CTC GGG GTA GCC TAC GCC CAG CTG      11998
Gly Val Thr Leu Gly Leu Val Val Leu Gly Val Ala Tyr Ala Gln Leu
    3985                3990                3995

GCC ATC CTG CTC GTG TCT TCC TGT GTG GAC TCC CTC TGG AGC GTG GCC      12046
Ala Ile Leu Leu Val Ser Ser Cys Val Asp Ser Leu Trp Ser Val Ala
4000                4005                4010                4015

CAG GCC CTG TTG GTG CTG TGC CCT GGG ACT GGG CTC TCT ACC CTG TGT      12094
Gln Ala Leu Leu Val Leu Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys
                    4020                4025                4030

CCT GCC GAG TCC TGG CAC CTG TCA CCC CTG CTG TGT GTG GGG CTC TGG      12142
Pro Ala Glu Ser Trp His Leu Ser Pro Leu Leu Cys Val Gly Leu Trp
            4035                4040                4045

GCA CTG CGG CTG TGG GGC GCC CTA CGG CTG GGG GCT GTT ATT CTC CGC      12190
Ala Leu Arg Leu Trp Gly Ala Leu Arg Leu Gly Ala Val Ile Leu Arg
        4050                4055                4060

TGG CGC TAC CAC GCC TTG CGT GGA GAG CTG TAC CGG CCG GCC TGG GAG      12238
Trp Arg Tyr His Ala Leu Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu
    4065                4070                4075
```

```
CCC CAG GAC TAC GAG ATG GTG GAG TTG TTC CTG CGC AGG CTG CGC CTC   12286
Pro Gln Asp Tyr Glu Met Val Glu Leu Phe Leu Arg Arg Leu Arg Leu
4080                4085                4090                4095

TGG ATG GGC CTC AGC AAG GTC AAG GAG TTC CGC CAC AAA GTC CGC TTT   12334
Trp Met Gly Leu Ser Lys Val Lys Glu Phe Arg His Lys Val Arg Phe
                    4100                4105                4110

GAA GGG ATG GAG CCG CTG CCC TCT CGC TCC TCC AGG GGC TCC AAG GTA   12382
Glu Gly Met Glu Pro Leu Pro Ser Arg Ser Ser Arg Gly Ser Lys Val
            4115                4120                4125

TCC CCG GAT GTG CCC CCA CCC AGC GCT GGC TCC GAT GCC TCG CAC CCC   12430
Ser Pro Asp Val Pro Pro Pro Ser Ala Gly Ser Asp Ala Ser His Pro
        4130                4135                4140

TCC ACC TCC TCC AGC CAG CTG GAT GGG CTG AGC GTG AGC CTG GGC CGG   12478
Ser Thr Ser Ser Ser Gln Leu Asp Gly Leu Ser Val Ser Leu Gly Arg
    4145                4150                4155

CTG GGG ACA AGG TGT GAG CCT GAG CCC TCC CGC CTC CAA GCC GTG TTC   12526
Leu Gly Thr Arg Cys Glu Pro Glu Pro Ser Arg Leu Gln Ala Val Phe
4160                4165                4170                4175

GAG GCC CTG CTC ACC CAG TTT GAC CGA CTC AAC CAG GCC ACA GAG GAC   12574
Glu Ala Leu Leu Thr Gln Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp
                    4180                4185                4190

GTC TAC CAG CTG GAG CAG CAG CTG CAC AGC CTG CAA GGC CGC AGG AGC   12622
Val Tyr Gln Leu Glu Gln Gln Leu His Ser Leu Gln Gly Arg Arg Ser
            4195                4200                4205

AGC CGG GCG CCC GCC GGA TCT TCC CGT GGC CCA TCC CCG GGC CTG CGG   12670
Ser Arg Ala Pro Ala Gly Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg
        4210                4215                4220

CCA GCA CTG CCC AGC CGC CTT GCC CGG GCC AGT CGG GGT GTG GAC CTG   12718
Pro Ala Leu Pro Ser Arg Leu Ala Arg Ala Ser Arg Gly Val Asp Leu
    4225                4230                4235

GCC ACT GGC CCC AGC AGG ACA CCT TCG GGC CAA GAA CAA GGT CCA CCC   12766
Ala Thr Gly Pro Ser Arg Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro
4240                4245                4250                4255

CAG CAG CAC TTA GTC CTC CTT CCT GGC GGG GGT GGG CCG TGG AGT CGG   12814
Gln Gln His Leu Val Leu Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg
                    4260                4265                4270

AGT GGA CAC CGC TCA GTA TTA CTT TCT GCC GCT GTC AAG GCC GAG GGC   12862
Ser Gly His Arg Ser Val Leu Leu Ser Ala Ala Val Lys Ala Glu Gly
            4275                4280                4285

CAG GCA GAA TGG CTG CAC GTA GGT TCC CCA GAG AGC AGG CAG GGG CAT   12910
Gln Ala Glu Trp Leu His Val Gly Ser Pro Glu Ser Arg Gln Gly His
        4290                4295                4300

CTG TCT GTC TGT GGG CTT CAG CAC TTT AAA GAG GCT GTG TGG CCA ACC   12958
Leu Ser Val Cys Gly Leu Gln His Phe Lys Glu Ala Val Trp Pro Thr
    4305                4310                4315
```

```
AGG ACC CAG GGT CCC CTC CCC AGC TCC CTT GGG AAG GAC ACA GCA GTA    13006
Arg Thr Gln Gly Pro Leu Pro Ser Ser Leu Gly Lys Asp Thr Ala Val
4320                4325                4330                4335

TTG GAC GGT TTC TAGCCTCTGA GATGCTAATT TATTTCCCCG AGTCCTCAGG      13058
Leu Asp Gly Phe

TACAGCGGGC TGTGCCCGGC CCCACCCCCT GGGCAGATGT CCCCCACTGC TAAGGCTGCT  13118

GGCTTCAGGG AGGGTTAGCC TGCACCGCCG CCACCCTGCC CCTAAGTTAT TACCTCTCCA  13178

GTTCCTACCG TACTCCCTGC ACCGTCTCAC TGTGTGTCTC GTGTCAGTAA TTTATATGGT  13238

GTTAAAATGT GTATATTTTT GTATGTCACT ATTTTCACTA GGGCTGAGGG GCCTGCGCCC  13298

AGAGCTGGCC TCCCCCAACA CCTGCTGCGC TTGGTAGGTG TGGTGGCGTT ATGGCAGCCC  13358

GGCTGCTGCT TGGATGCGAG CTTGGCCTTG GGCCGGTGCT GGGGGCACAG CTGTCTGCCA  13418

GGCACTCTCA TCACCCCAGA GGCCTTGTCA TCCTCCCTTG CCCCAGGCCA GGTAGCAAGA  13478

GAGCAGCGCC CAGGCCTGCT GGCATCAGGT CTGGGCAAGT AGCAGGACTA GGCATGTCAG  13538

AGGACCCCAG GGTGGTTAGA GGAAAAGACT CCTCCTGGGG GCTGGCTCCC AGGGTGGAGG  13598

AAGGTGACTG TGTGTGTGTG TGTGTGCGCG CGCGACGCGC GAGTGTGCTG TATGGCCCAG  13658

GCAGCCTCAA GGCCCTCGGA GCTGGCTGTG CCTGCTTCTG TGTACCACTT CTGTGGGCAT  13718

GGCCGCTTCT AGAGCCTCGA CACCCCCCCA ACCCCCGCAC CAAGCAGACA AAGTCAATAA  13778

AAGAGCTGTC TGACTGCAAA AAAAAAAAA                                    13807
```

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Gly Ala Ala Cys Arg Val Asn Cys Ser Gly Arg Gly Leu Arg Thr Leu
 1               5                  10                  15

Gly Pro Ala Leu Arg Ile Pro Ala Asp Ala Thr Ala Leu Asp Val Ser
            20                  25                  30

His Asn Leu Leu Arg Ala Leu Asp Val Gly Leu Leu Ala Asn Leu Ser
        35                  40                  45

Ala Leu Ala Glu Leu Asp Ile Ser Asn Asn Lys Ile Ser Thr Leu Glu
    50                  55                  60

Glu Gly Ile Phe Ala Asn Leu Phe Asn Leu Ser Glu Ile Asn Leu Ser
65                  70                  75                  80

Gly Asn Pro Phe Glu Cys Asp Cys Gly Leu Ala Trp Leu Pro Arg Trp
                85                  90                  95

Ala Glu Glu Gln Gln Val Arg Val Val Gln Pro Glu Ala Ala Thr Cys
            100                 105                 110

Ala Gly Pro Gly Ser Leu Ala Gly Gln Pro Leu Leu Gly Ile Pro Leu
        115                 120                 125
```

Leu Asp Ser Gly Cys Gly Glu Glu Tyr Val Ala Cys Leu Pro Asp Asn
130 135 140

Ser Ser Gly Thr Val Ala Ala Val Ser Phe Ser Ala Ala His Glu Gly
145 150 155 160

Leu Leu Gln Pro Glu Ala Cys Ser Ala Phe Cys Phe Ser Thr Gly Gln
165 170 175

Gly Leu Ala Ala Leu Ser Glu Gln Gly Trp Cys Leu Cys Gly Ala Ala
180 185 190

Gln Pro Ser Ser Ala Ser Phe Ala Cys Leu Ser Leu Cys Ser Gly Pro
195 200 205

Pro Pro Pro Pro Ala Pro Thr Cys Arg Gly Pro Thr Leu Leu Gln His
210 215 220

Val Phe Pro Ala Ser Pro Gly Ala Thr Leu Val Gly Pro His Gly Pro
225 230 235 240

Leu Ala Ser Gly Gln Leu Ala Ala Phe His Ile Ala Ala Pro Leu Pro
245 250 255

Val Thr Ala Thr Arg Trp Asp Phe Gly Asp Gly Ser Ala Glu Val Asp
260 265 270

Ala Ala Gly Pro Ala Ala Ser His Arg Tyr Val Leu Pro Gly Arg Tyr
275 280 285

His Val Thr Ala Val Leu Ala Leu Gly Ala Gly Ser Ala Leu Leu Gly
290 295 300

Thr Asp Val Gln Val Glu Ala Ala Pro Ala Ala Leu Glu Leu Val Cys
305 310 315 320

Pro Ser Ser Val Gln Ser Asp Glu Ser Leu Asp Leu Ser Ile Gln Asn
325 330 335

Arg Gly Gly Ser Gly Leu Glu Ala Ala Tyr Ser Ile Val Ala Leu Gly
340 345 350

Glu Glu Pro Ala Arg Ala Val His Pro Leu Cys Pro Ser Asp Thr Glu
355 360 365

Ile Phe Pro Gly Asn Gly His Cys Tyr Arg Leu Val Val Glu Lys Ala
370 375 380

Ala Trp Leu Gln Ala Gln Glu Gln Cys Gln Ala Trp Ala Gly Ala Ala
385 390 395 400

Leu Ala Met Val Asp Ser Pro Ala Val Gln Arg Phe Leu Val Ser Arg
405 410 415

Val Thr Arg Ser Leu Asp Val Trp Ile Gly Phe Ser Thr Val Gln Gly
420 425 430

Val Glu Val Gly Pro Ala Pro Gln Gly Glu Ala Phe Ser Leu Glu Ser
435 440 445

Cys Gln Asn Trp Leu Pro Gly Glu Pro His Pro Ala Thr Ala Glu His
450 455 460

```
Cys Val Arg Leu Gly Pro Thr Gly Trp Cys Asn Thr Asp Leu Cys Ser
465                 470                 475                 480

Ala Pro His Ser Tyr Val Cys Glu Leu Gln Pro Gly Gly Pro Val Gln
                    485                 490                 495

Asp Ala Glu Asn Leu Leu Val Gly Ala Pro Ser Gly Asp Leu Gln Gly
            500                 505                 510

Pro Leu Thr Pro Leu Ala Gln Gln Asp Gly Leu Ser Ala Pro His Glu
        515                 520                 525

Pro Val Glu Val Met Val Phe Pro Gly Leu Arg Leu Ser Arg Glu Ala
    530                 535                 540

Phe Leu Thr Thr Ala Glu Phe Gly Thr Gln Glu Leu Arg Arg Pro Ala
545                 550                 555                 560

Gln Leu Arg Leu Gln Val Tyr Arg Leu Leu Ser Thr Ala Gly Thr Pro
                565                 570                 575

Glu Asn Gly Ser Glu Pro Glu Ser Arg Ser Pro Asp Asn Arg Thr Gln
            580                 585                 590

Leu Ala Pro Ala Cys Met Pro Gly Gly Arg Trp Cys Pro Gly Ala Asn
        595                 600                 605

Ile Cys Leu Pro Leu Asp Ala Ser Cys His Pro Gln Ala Cys Ala Asn
    610                 615                 620

Gly Cys Thr Ser Gly Pro Gly Leu Pro Gly Ala Pro Tyr Ala Leu Trp
625                 630                 635                 640

Arg Glu Phe Leu Phe Ser Val Ala Ala Gly Pro Pro Ala Gln Tyr Ser
                645                 650                 655

Val Thr Leu His Gly Gln Asp Val Leu Met Leu Pro Gly Asp Leu Val
            660                 665                 670

Gly Leu Gln His Asp Ala Gly Pro Gly Ala Leu Leu His Cys Ser Pro
        675                 680                 685

Ala Pro Gly His Pro Gly Pro Gln Ala Pro Tyr Leu Ser Ala Asn Ala
    690                 695                 700

Ser Ser Trp Leu Pro His Leu Pro Ala Gln Leu Glu Gly Thr Trp Ala
705                 710                 715                 720

Cys Pro Ala Cys Ala Leu Arg Leu Leu Ala Ala Thr Glu Gln Leu Thr
                725                 730                 735

Val Leu Leu Gly Leu Arg Pro Asn Pro Gly Leu Arg Met Pro Gly Arg
            740                 745                 750

Tyr Glu Val Arg Ala Glu Val Gly Asn Gly Val Ser Arg His Asn Leu
        755                 760                 765

Ser Cys Ser Phe Asp Val Val Ser Pro Val Ala Gly Leu Arg Val Ile
    770                 775                 780

Tyr Pro Ala Pro Arg Asp Gly Arg Leu Tyr Val Pro Thr Asn Gly Ser
785                 790                 795                 800
```

Ala Leu Val Leu Gln Val Asp Ser Gly Ala Asn Ala Thr Ala Thr Ala
805 810 815

Arg Trp Pro Gly Gly Ser Val Ser Ala Arg Phe Glu Asn Val Cys Pro
820 825 830

Ala Leu Val Ala Thr Phe Val Pro Gly Cys Pro Trp Glu Thr Asn Asp
835 840 845

Thr Leu Phe Ser Val Val Ala Leu Pro Trp Leu Ser Glu Gly Glu His
850 855 860

Val Val Asp Val Val Val Glu Asn Ser Ala Ser Arg Ala Asn Leu Ser
865 870 875 880

Leu Arg Val Thr Ala Glu Glu Pro Ile Cys Gly Leu Arg Ala Thr Pro
885 890 895

Ser Pro Glu Ala Arg Val Leu Gln Gly Val Leu Val Arg Tyr Ser Pro
900 905 910

Val Val Glu Ala Gly Ser Asp Met Val Phe Arg Trp Thr Ile Asn Asp
915 920 925

Lys Gln Ser Leu Thr Phe Gln Asn Val Val Phe Asn Val Ile Tyr Gln
930 935 940

Ser Ala Ala Val Phe Lys Leu Ser Leu Thr Ala Ser Asn His Val Ser
945 950 955 960

Asn Val Thr Val Asn Tyr Asn Val Thr Val Glu Arg Met Asn Arg Met
965 970 975

Gln Gly Leu Gln Val Ser Thr Val Pro Ala Val Leu Ser Pro Asn Ala
980 985 990

Thr Leu Val Leu Thr Gly Gly Val Leu Val Asp Ser Ala Val Glu Val
995 1000 1005

Ala Phe Leu Trp Asn Phe Gly Asp Gly Glu Gln Ala Leu His Gln Phe
1010 1015 1020

Gln Pro Pro Tyr Asn Glu Ser Phe Pro Val Pro Asp Pro Ser Val Ala
1025 1030 1035 1040

Gln Val Leu Val Glu His Asn Val Met His Thr Tyr Ala Ala Pro Gly
1045 1050 1055

Glu Tyr Leu Leu Thr Val Leu Ala Ser Asn Ala Phe Glu Asn Leu Thr
1060 1065 1070

Gln Gln Val Pro Val Ser Val Arg Ala Ser Leu Pro Ser Val Ala Val
1075 1080 1085

Gly Val Ser Asp Gly Val Leu Val Ala Gly Arg Pro Val Thr Phe Tyr
1090 1095 1100

Pro His Pro Leu Pro Ser Pro Gly Gly Val Leu Tyr Thr Trp Asp Phe
1105 1110 1115 1120

Gly Asp Gly Ser Pro Val Leu Thr Gln Ser Gln Pro Ala Ala Asn His
1125 1130 1135

Thr Tyr Ala Ser Arg Gly Thr Tyr His Val Arg Leu Glu Val Asn Asn
1140 1145 1150

Thr Val Ser Gly Ala Ala Ala Gln Ala Asp Val Arg Val Phe Glu Glu
1155 1160 1165

Leu Arg Gly Leu Ser Val Asp Met Ser Leu Ala Val Glu Gln Gly Ala
1170 1175 1180

Pro Val Val Val Ser Ala Ala Val Gln Thr Gly Asp Asn Ile Thr Trp
1185 1190 1195 1200

Thr Phe Asp Met Gly Asp Gly Thr Val Leu Ser Gly Pro Glu Ala Thr
1205 1210 1215

Val Glu His Val Tyr Leu Arg Ala Gln Asn Cys Thr Val Thr Val Gly
1220 1225 1230

Ala Ala Ser Pro Ala Gly His Leu Ala Arg Ser Leu His Val Leu Val
1235 1240 1245

Phe Val Leu Glu Val Leu Arg Val Glu Pro Ala Ala Cys Ile Pro Thr
1250 1255 1260

Gln Pro Asp Ala Arg Leu Thr Ala Tyr Val Thr Gly Asn Pro Ala His
1265 1270 1275 1280

Tyr Leu Phe Asp Trp Thr Phe Gly Asp Gly Ser Ser Asn Thr Thr Val
1285 1290 1295

Arg Gly Cys Pro Thr Val Thr His Asn Phe Thr Arg Ser Gly Thr Phe
1300 1305 1310

Pro Leu Ala Leu Val Leu Ser Ser Arg Val Asn Arg Ala His Tyr Phe
1315 1320 1325

Thr Ser Ile Cys Val Glu Pro Glu Val Gly Asn Val Thr Leu Gln Pro
1330 1335 1340

Glu Arg Gln Phe Val Gln Leu Gly Asp Glu Ala Trp Leu Val Ala Cys
1345 1350 1355 1360

Ala Trp Pro Pro Phe Pro Tyr Arg Tyr Thr Trp Asp Phe Gly Thr Glu
1365 1370 1375

Glu Ala Ala Pro Thr Arg Ala Arg Gly Pro Glu Val Thr Phe Ile Tyr
1380 1385 1390

Arg Asp Pro Gly Ser Tyr Leu Val Thr Val Thr Ala Ser Asn Asn Ile
1395 1400 1405

Ser Ala Ala Asn Asp Ser Ala Leu Val Glu Val Gln Glu Pro Val Leu
1410 1415 1420

Val Thr Ser Ile Lys Val Asn Gly Ser Leu Gly Leu Glu Leu Gln Gln
1425 1430 1435 1440

Pro Tyr Leu Phe Ser Ala Val Gly Arg Gly Arg Pro Ala Ser Tyr Leu
1445 1450 1455

Trp Asp Leu Gly Asp Gly Gly Trp Leu Glu Gly Pro Glu Val Thr His
1460 1465 1470

Ala Tyr Asn Ser Thr Gly Asp Phe Thr Val Arg Val Ala Gly Trp Asn
1475 1480 1485

Glu Val Ser Arg Ser Glu Ala Trp Leu Asn Val Thr Val Lys Arg Arg
1490 1495 1500

Val Arg Gly Leu Val Val Asn Ala Ser Arg Thr Val Val Pro Leu Asn
1505 1510 1515 1520

Gly Ser Val Ser Phe Ser Thr Ser Leu Glu Ala Gly Ser Asp Val Arg
1525 1530 1535

Tyr Ser Trp Val Leu Cys Asp Arg Cys Thr Pro Ile Pro Gly Gly Pro
1540 1545 1550

Thr Ile Ser Tyr Thr Phe Arg Ser Val Gly Thr Phe Asn Ile Ile Val
1555 1560 1565

Thr Ala Glu Asn Glu Val Gly Ser Ala Gln Asp Ser Ile Phe Val Tyr
1570 1575 1580

Val Leu Gln Leu Ile Glu Gly Leu Gln Val Val Gly Gly Gly Arg Tyr
1585 1590 1595 1600

Phe Pro Thr Asn His Thr Val Gln Leu Gln Ala Val Val Arg Asp Gly
1605 1610 1615

Thr Asn Val Ser Tyr Ser Trp Thr Ala Trp Arg Asp Arg Gly Pro Ala
1620 1625 1630

Leu Ala Gly Ser Gly Lys Gly Phe Ser Leu Thr Val Leu Glu Ala Gly
1635 1640 1645

Thr Tyr His Val Gln Leu Arg Ala Thr Asn Met Leu Gly Ser Ala Trp
1650 1655 1660

Ala Asp Cys Thr Met Asp Phe Val Glu Pro Val Gly Trp Leu Met Val
1665 1670 1675 1680

Thr Ala Ser Pro Asn Pro Ala Ala Val Asn Thr Ser Val Thr Leu Ser
1685 1690 1695

Ala Glu Leu Ala Gly Gly Ser Gly Val Val Tyr Thr Trp Ser Leu Glu
1700 1705 1710

Glu Gly Leu Ser Trp Glu Thr Ser Glu Pro Phe Thr Thr His Ser Phe
1715 1720 1725

Pro Thr Pro Gly Leu His Leu Val Thr Met Thr Ala Gly Asn Pro Leu
1730 1735 1740

Gly Ser Ala Asn Ala Thr Val Glu Val Asp Val Gln Val Pro Val Ser
1745 1750 1755 1760

Gly Leu Ser Ile Arg Ala Ser Glu Pro Gly Gly Ser Phe Val Ala Ala
1765 1770 1775

Gly Ser Ser Val Pro Phe Trp Gly Gln Leu Ala Thr Gly Thr Asn Val
1780 1785 1790

Ser Trp Cys Trp Ala Val Pro Gly Gly Ser Ser Lys Arg Gly Pro His
1795 1800 1805

Val Thr Met Val Phe Pro Asp Ala Gly Thr Phe Ser Ile Arg Leu Asn
1810 1815 1820

Ala Ser Asn Ala Val Ser Trp Val Ser Ala Thr Tyr Asn Leu Thr Ala
1825 1830 1835 1840

Glu Glu Pro Ile Val Gly Leu Val Leu Trp Ala Ser Ser Lys Val Val
1845 1850 1855

Ala Pro Gly Gln Leu Val His Phe Gln Ile Leu Leu Ala Ala Gly Ser
1860 1865 1870

Ala Val Thr Phe Arg Leu Gln Val Gly Gly Ala Asn Pro Glu Val Leu
1875 1880 1885

Pro Gly Pro Arg Phe Ser His Ser Phe Pro Arg Val Gly Asp His Val
1890 1895 1900

Val Ser Val Arg Gly Lys Asn His Val Ser Trp Ala Gln Ala Gln Val
1905 1910 1915 1920

Arg Ile Val Val Leu Glu Ala Val Ser Gly Leu Gln Met Pro Asn Cys
1925 1930 1935

Cys Glu Pro Gly Ile Ala Thr Gly Thr Glu Arg Asn Phe Thr Ala Arg
1940 1945 1950

Val Gln Arg Gly Ser Arg Val Ala Tyr Ala Trp Tyr Phe Ser Leu Gln
1955 1960 1965

Lys Val Gln Gly Asp Ser Leu Val Ile Leu Ser Gly Arg Asp Val Thr
1970 1975 1980

Tyr Thr Pro Val Ala Ala Gly Leu Leu Glu Ile Gln Val Arg Ala Phe
1985 1990 1995 2000

Asn Ala Leu Gly Ser Glu Asn Arg Thr Leu Val Leu Glu Val Gln Asp
2005 2010 2015

Ala Val Gln Tyr Val Ala Leu Gln Ser Gly Pro Cys Phe Thr Asn Arg
2020 2025 2030

Ser Ala Gln Phe Glu Ala Ala Thr Ser Pro Ser Pro Arg Arg Val Ala
2035 2040 2045

Tyr His Trp Asp Phe Gly Asp Gly Ser Pro Gly Gln Asp Thr Asp Glu
2050 2055 2060

Pro Arg Ala Glu His Ser Tyr Leu Arg Pro Gly Asp Tyr Arg Val Gln
2065 2070 2075 2080

Val Asn Ala Ser Asn Leu Val Ser Phe Phe Val Ala Gln Ala Thr Val
2085 2090 2095

Thr Val Gln Val Leu Ala Cys Arg Glu Pro Glu Val Asp Val Val Leu
2100 2105 2110

Pro Leu Gln Val Leu Met Arg Arg Ser Gln Arg Asn Tyr Leu Glu Ala
2115 2120 2125

His Val Asp Leu Arg Asp Cys Val Thr Tyr Gln Thr Glu Tyr Arg Trp
2130 2135 2140

Glu Val Tyr Arg Thr Ala Ser Cys Gln Arg Pro Gly Arg Pro Ala Arg
2145 2150 2155 2160

Val Ala Leu Pro Gly Val Asp Val Ser Arg Pro Arg Leu Val Leu Pro
2165 2170 2175

Arg Leu Ala Leu Pro Val Gly His Tyr Cys Phe Val Phe Val Val Ser
2180 2185 2190

Phe Gly Asp Thr Pro Leu Thr Gln Ser Ile Gln Ala Asn Val Thr Val
2195 2200 2205

Ala Pro Glu Arg Leu Val Pro Ile Ile Glu Gly Gly Ser Tyr Arg Val
2210 2215 2220

Trp Ser Asp Thr Arg Asp Leu Val Leu Asp Gly Ser Glu Ser Tyr Asp
2225 2230 2235 2240

Pro Asn Leu Glu Asp Gly Asp Gln Thr Pro Leu Ser Phe His Trp Ala
2245 2250 2255

Cys Val Ala Ser Thr Gln Arg Glu Ala Gly Gly Cys Ala Leu Asn Phe
2260 2265 2270

Gly Pro Arg Gly Ser Ser Thr Val Thr Ile Pro Arg Glu Arg Leu Ala
2275 2280 2285

Ala Gly Val Glu Tyr Thr Phe Ser Leu Thr Val Trp Lys Ala Gly Arg
2290 2295 2300

Lys Glu Glu Ala Thr Asn Gln Thr Val Leu Ile Arg Ser Gly Arg Val
2305 2310 2315 2320

Pro Ile Val Ser Leu Glu Cys Val Ser Cys Lys Ala Gln Ala Val Tyr
2325 2330 2335

Glu Val Ser Arg Ser Ser Tyr Val Tyr Leu Glu Gly Arg Cys Leu Asn
2340 2345 2350

Cys Ser Ser Gly Ser Lys Arg Gly Arg Trp Ala Ala Arg Thr Phe Ser
2355 2360 2365

Asn Lys Thr Leu Val Leu Asp Glu Thr Thr Thr Ser Thr Gly Ser Ala
2370 2375 2380

Gly Met Arg Leu Val Leu Arg Arg Gly Val Leu Arg Asp Gly Glu Gly
2385 2390 2395 2400

Tyr Thr Phe Thr Leu Thr Val Leu Gly Arg Ser Gly Glu Glu Glu Gly
2405 2410 2415

Cys Ala Ser Ile Arg Leu Ser Pro Asn Arg Pro Pro Leu Gly Gly Ser
2420 2425 2430

Cys Arg Leu Phe Pro Leu Gly Ala Val His Ala Leu Thr Thr Lys Val
2435 2440 2445

His Phe Glu Cys Thr Gly Trp His Asp Ala Glu Asp Ala Gly Ala Pro
2450 2455 2460

Leu Val Tyr Ala Leu Leu Leu Arg Arg Cys Arg Gln Gly His Cys Glu
2465 2470 2475 2480

Glu Phe Cys Val Tyr Lys Gly Ser Leu Ser Ser Tyr Gly Ala Val Leu
2485 2490 2495

Pro Pro Gly Phe Arg Pro His Phe Glu Val Gly Leu Ala Val Val Val
2500 2505 2510

Gln Asp Gln Leu Gly Ala Ala Val Val Ala Leu Asn Arg Ser Leu Ala
2515 2520 2525

Ile Thr Leu Pro Glu Pro Asn Gly Ser Ala Thr Gly Leu Thr Val Trp
2530 2535 2540

Leu His Gly Leu Thr Ala Ser Val Leu Pro Gly Leu Leu Arg Gln Ala
2545 2550 2555 2560

Asp Pro Gln His Val Ile Glu Tyr Ser Leu Ala Leu Val Thr Val Leu
2565 2570 2575

Asn Glu Tyr Glu Arg Ala Leu Asp Val Ala Ala Glu Pro Lys His Glu
2580 2585 2590

Arg Gln His Arg Ala Gln Ile Arg Lys Asn Ile Thr Glu Thr Leu Val
2595 2600 2605

Ser Leu Arg Val His Thr Val Asp Asp Ile Gln Gln Ile Ala Ala Ala
2610 2615 2620

Leu Ala Gln Cys Met Gly Pro Ser Arg Glu Leu Val Cys Arg Ser Cys
2625 2630 2635 2640

Leu Lys Gln Thr Leu His Lys Leu Glu Ala Met Met Leu Ile Leu Gln
2645 2650 2655

Ala Glu Thr Thr Ala Gly Thr Val Thr Pro Thr Ala Ile Gly Asp Ser
2660 2665 2670

Ile Leu Asn Ile Thr Gly Asp Leu Ile His Leu Ala Ser Ser Asp Val
2675 2680 2685

Arg Ala Pro Gln Pro Ser Glu Leu Gly Ala Glu Ser Pro Ser Arg Met
2690 2695 2700

Val Ala Ser Gln Ala Tyr Asn Leu Thr Ser Ala Leu Met Arg Ile Leu
2705 2710 2715 2720

Met Arg Ser Arg Val Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly Glu
2725 2730 2735

Glu Ile Val Ala Gln Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu Cys
2740 2745 2750

Tyr Gly Gly Ala Pro Gly Pro Gly Cys His Phe Ser Ile Pro Glu Ala
2755 2760 2765

Phe Ser Gly Ala Leu Ala Asn Leu Ser Asp Val Val Gln Leu Ile Phe
2770 2775 2780

Leu Val Asp Ser Asn Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr Thr
2785 2790 2795 2800

Val Ser Thr Lys Val Ala Ser Met Ala Phe Gln Thr Gln Ala Gly Ala
2805 2810 2815

Gln Ile Pro Ile Glu Arg Leu Ala Ser Glu Arg Ala Ile Thr Val Lys
2820 2825 2830

Val Pro Asn Asn Ser Asp Trp Ala Ala Arg Gly His Arg Ser Ser Ala
2835 2840 2845

Asn Ser Ala Asn Ser Val Val Val Gln Pro Gln Ala Ser Val Gly Ala
2850 2855 2860

Val Val Thr Leu Asp Ser Ser Asn Pro Ala Ala Gly Leu His Leu Gln
2865 2870 2875 2880

Leu Asn Tyr Thr Leu Leu Asp Gly His Tyr Leu Ser Glu Glu Pro Glu
2885 2890 2895

Pro Tyr Leu Ala Val Tyr Leu His Ser Glu Pro Arg Pro Asn Glu His
2900 2905 2910

Asn Cys Ser Ala Ser Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly Ala
2915 2920 2925

Asp His Arg Pro Tyr Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp Pro
2930 2935 2940

Ala Gly Ser Tyr His Leu Asn Leu Ser Ser His Phe Arg Trp Ser Ala
2945 2950 2955 2960

Leu Gln Val Ser Val Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe Ser
2965 2970 2975

Glu Glu Asp Met Val Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu Glu
2980 2985 2990

Thr Ser Pro Arg Gln Ala Val Cys Leu Thr Arg His Leu Thr Ala Phe
2995 3000 3005

Gly Ala Ser Leu Phe Val Pro Pro Ser His Val Arg Phe Val Phe Pro
3010 3015 3020

Glu Pro Thr Ala Asp Val Asn Tyr Ile Val Met Leu Thr Cys Ala Val
3025 3030 3035 3040

Cys Leu Val Thr Tyr Met Val Met Ala Ala Ile Leu His Lys Leu Asp
3045 3050 3055

Gln Leu Asp Ala Ser Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln Arg
3060 3065 3070

Gly Arg Phe Lys Tyr Glu Ile Leu Val Lys Thr Gly Trp Gly Arg Gly
3075 3080 3085

Ser Gly Thr Thr Ala His Val Gly Ile Met Leu Tyr Gly Val Asp Ser
3090 3095 3100

Arg Ser Gly His Arg His Leu Asp Gly Asp Arg Ala Phe His Arg Asn
3105 3110 3115 3120

Ser Leu Asp Ile Phe Arg Ile Ala Thr Pro His Ser Leu Gly Ser Val
3125 3130 3135

Trp Lys Ile Arg Val Trp His Asp Asn Lys Gly Leu Ser Pro Ala Trp
3140 3145 3150

Phe Leu Gln His Val Ile Val Arg Asp Leu Gln Thr Ala Arg Ser Ala
3155 3160 3165

Phe Phe Leu Val Asn Asp Trp Leu Ser Val Glu Thr Glu Ala Asn Gly
3170 3175 3180

Gly Leu Val Glu Lys Glu Val Leu Ala Ala Ser Asp Ala Ala Leu Leu
3185 3190 3195 3200

Arg Phe Arg Arg Leu Leu Val Ala Glu Leu Gln Arg Gly Phe Phe Asp
3205 3210 3215

Lys His Ile Trp Leu Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg Phe
3220 3225 3230

Thr Arg Ile Gln Arg Ala Thr Cys Cys Val Leu Leu Ile Cys Leu Phe
3235 3240 3245

Leu Gly Ala Asn Ala Val Trp Tyr Gly Ala Val Gly Asp Ser Ala Tyr
3250 3255 3260

Ser Thr Gly His Val Ser Arg Leu Ser Pro Leu Ser Val Asp Thr Val
3265 3270 3275 3280

Ala Val Gly Leu Val Ser Ser Val Val Val Tyr Pro Val Tyr Leu Ala
3285 3290 3295

Ile Leu Phe Leu Phe Arg Met Ser Arg Ser Lys Val Ala Gly Ser Pro
3300 3305 3310

Ser Pro Thr Pro Ala Gly Gln Gln Val Leu Asp Ile Asp Ser Cys Leu
3315 3320 3325

Asp Ser Ser Val Leu Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu His
3330 3335 3340

Ala Glu Ala Phe Val Gly Gln Met Lys Ser Asp Leu Phe Leu Asp Asp
3345 3350 3355 3360

Ser Lys Ser Leu Val Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser Trp
3365 3370 3375

Pro Asp Leu Leu Ser Asp Pro Ser Ile Val Gly Ser Asn Leu Arg Gln
3380 3385 3390

Leu Ala Arg Gly Gln Ala Gly His Gly Leu Gly Pro Glu Glu Asp Gly
3395 3400 3405

Phe Ser Leu Ala Ser Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala Ser
3410 3415 3420

Asp Glu Asp Leu Ile Gln Gln Val Leu Ala Glu Gly Val Ser Ser Pro
3425 3430 3435 3440

Ala Pro Thr Gln Asp Thr His Met Glu Thr Asp Leu Leu Ser Ser Leu
3445 3450 3455

Ser Ser Thr Pro Gly Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg Leu
3460 3465 3470

Gly Glu Leu Gly Pro Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro Gln
3475 3480 3485

Ala Ala Arg Leu Ser Arg Thr Gly Leu Val Glu Gly Leu Arg Lys Arg
3490 3495 3500

Leu Leu Pro Ala Trp Cys Ala Ser Leu Ala His Gly Leu Ser Leu Leu
3505 3510 3515 3520

Leu Val Ala Val Ala Val Ala Val Ser Gly Trp Val Gly Ala Ser Phe
3525 3530 3535

Pro Pro Gly Val Ser Val Ala Trp Leu Leu Ser Ser Ser Ala Ser Phe
3540 3545 3550

Leu Ala Ser Phe Leu Gly Trp Glu Pro Leu Lys Val Leu Leu Glu Ala
3555 3560 3565

Leu Tyr Phe Ser Leu Val Ala Lys Arg Leu His Pro Asp Glu Asp Asp
3570 3575 3580

Thr Leu Val Glu Ser Pro Ala Val Thr Pro Val Ser Ala Arg Val Pro
3585 3590 3595 3600

Arg Val Arg Pro Pro His Gly Phe Ala Leu Phe Leu Ala Lys Glu Glu
3605 3610 3615

Ala Arg Lys Val Lys Arg Leu His Gly Met Leu Arg Ser Leu Leu Val
3620 3625 3630

Tyr Met Leu Phe Leu Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp Ala
3635 3640 3645

Ser Cys His Gly His Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln Glu
3650 3655 3660

Leu His Ser Arg Ala Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu Trp
3665 3670 3675 3680

Pro Trp Met Ala His Val Leu Leu Pro Tyr Val His Gly Asn Gln Ser
3685 3690 3695

Ser Pro Glu Leu Gly Pro Pro Arg Leu Arg Gln Val Arg Leu Gln Glu
3700 3705 3710

Ala Leu Tyr Pro Asp Pro Pro Gly Pro Arg Val His Thr Cys Ser Ala
3715 3720 3725

Ala Gly Gly Phe Ser Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser Pro
3730 3735 3740

His Asn Gly Ser Gly Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu Gly
3745 3750 3755 3760

Ala Trp Ser Trp Gly Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr Val
3765 3770 3775

Gln Glu Leu Gly Leu Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg Phe
3780 3785 3790

Leu Gln Leu His Asn Trp Leu Asp Asn Arg Ser Arg Ala Val Phe Leu
3795 3800 3805

Glu Leu Thr Arg Tyr Ser Pro Ala Val Gly Leu His Ala Ala Val Thr
3810 3815 3820

Leu Arg Leu Glu Phe Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu Ser
3825 3830 3835 3840

Val Arg Pro Phe Ala Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu Pro
3845 3850 3855

Leu Leu Thr Ser Val Cys Leu Leu Leu Phe Ala Val His Phe Ala Val
3860 3865 3870

Ala Glu Ala Arg Thr Trp His Arg Glu Gly Arg Trp Arg Val Leu Arg
3875 3880 3885

Leu Gly Ala Trp Ala Arg Trp Leu Leu Val Ala Leu Thr Ala Ala Thr
3890 3895 3900

Ala Leu Val Arg Leu Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp Thr
3905 3910 3915 3920

Arg Phe Val Arg Gly Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln Val
3925 3930 3935

Ala His Val Ser Ser Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu Phe
3940 3945 3950

Leu Leu Leu Val Lys Ala Ala Gln His Val Arg Phe Val Arg Gln Trp
3955 3960 3965

Ser Val Phe Gly Lys Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu Gly
3970 3975 3980

Val Thr Leu Gly Leu Val Val Leu Gly Val Ala Tyr Ala Gln Leu Ala
3985 3990 3995 4000

Ile Leu Leu Val Ser Ser Cys Val Asp Ser Leu Trp Ser Val Ala Gln
4005 4010 4015

Ala Leu Leu Val Leu Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys Pro
4020 4025 4030

Ala Glu Ser Trp His Leu Ser Pro Leu Leu Cys Val Gly Leu Trp Ala
4035 4040 4045

Leu Arg Leu Trp Gly Ala Leu Arg Leu Gly Ala Val Ile Leu Arg Trp
4050 4055 4060

Arg Tyr His Ala Leu Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu Pro
4065 4070 4075 4080

Gln Asp Tyr Glu Met Val Glu Leu Phe Leu Arg Arg Leu Arg Leu Trp
4085 4090 4095

Met Gly Leu Ser Lys Val Lys Glu Phe Arg His Lys Val Arg Phe Glu
4100 4105 4110

Gly Met Glu Pro Leu Pro Ser Arg Ser Ser Arg Gly Ser Lys Val Ser
4115 4120 4125

Pro Asp Val Pro Pro Pro Ser Ala Gly Ser Asp Ala Ser His Pro Ser
4130 4135 4140

Thr Ser Ser Ser Gln Leu Asp Gly Leu Ser Val Ser Leu Gly Arg Leu
4145 4150 4155 4160

Gly Thr Arg Cys Glu Pro Glu Pro Ser Arg Leu Gln Ala Val Phe Glu
4165 4170 4175

Ala Leu Leu Thr Gln Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp Val
4180 4185 4190

Tyr Gln Leu Glu Gln Gln Leu His Ser Leu Gln Gly Arg Arg Ser Ser
4195 4200 4205

Arg Ala Pro Ala Gly Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg Pro
4210 4215 4220

Ala Leu Pro Ser Arg Leu Ala Arg Ala Ser Arg Gly Val Asp Leu Ala
4225 4230 4235 4240

Thr Gly Pro Ser Arg Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro Gln
4245 4250 4255

Gln His Leu Val Leu Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg Ser
4260 4265 4270

Gly His Arg Ser Val Leu Leu Ser Ala Ala Val Lys Ala Glu Gly Gln
4275 4280 4285

Ala Glu Trp Leu His Val Gly Ser Pro Glu Ser Arg Gln Gly His Leu
4290 4295 4300

Ser Val Cys Gly Leu Gln His Phe Lys Glu Ala Val Trp Pro Thr Arg
4305 4310 4315 4320

Thr Gln Gly Pro Leu Pro Ser Ser Leu Gly Lys Asp Thr Ala Val Leu
4325 4330 4335

Asp Gly Phe

Figure 10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3: (Compare Figure 7)

```
CTC AAC GAG GAG CCC CTG ACG CTG GCG GGC GAG GAG ATC GTG GCC CAG      48
Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly Glu Glu Ile Val Ala Gln
4340                4345                4350                4355

GGC AAG CGC TCG GAC CCG CGG AGC CTG CTG TGC TAT GGC GGC GCC CCA      96
Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu Cys Tyr Gly Gly Ala Pro
                    4360                4365                4370

GGG CCT GGC TGC CAC TTC TCC ATC CCC GAG GCT TTC AGC GGG GCC CTG     144
Gly Pro Gly Cys His Phe Ser Ile Pro Glu Ala Phe Ser Gly Ala Leu
            4375                4380                4385

GCC AAC CTC AGT GAC GTG GTG CAG CTC ATC TTT CTG GTG GAC TCC AAT     192
Ala Asn Leu Ser Asp Val Val Gln Leu Ile Phe Leu Val Asp Ser Asn
        4390                4395                4400

CCC TTT CCC TTT GGC TAT ATC AGC AAC TAC ACC GTC TCC ACC AAG GTG     240
Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr Thr Val Ser Thr Lys Val
    4405                4410                4415

GCC TCG ATG GCA TTC CAG ACA CAG GCC GGC GCC CAG ATC CCC ATC GAG     288
Ala Ser Met Ala Phe Gln Thr Gln Ala Gly Ala Gln Ile Pro Ile Glu
4420                4425                4430                4435

CGG CTG GCC TCA GAG CGC GCC ATC ACC GTG AAG GTG CCC AAC AAC TCG     336
Arg Leu Ala Ser Glu Arg Ala Ile Thr Val Lys Val Pro Asn Asn Ser
                    4440                4445                4450

GAC TGG GCT GCC CGG GGC CAC CGC AGC TCC GCC AAC TCC GCC AAC TCC     384
Asp Trp Ala Ala Arg Gly His Arg Ser Ser Ala Asn Ser Ala Asn Ser
            4455                4460                4465

GTT GTG GTC CAG CCC CAG GCC TCC GTC GGT GCT GTG GTC ACC CTG GAC     432
Val Val Val Gln Pro Gln Ala Ser Val Gly Ala Val Val Thr Leu Asp
        4470                4475                4480

AGC AGC AAC CCT GCG GCC GGG CTG CAT CTG CAG CTC AAC TAT ACG CTG     480
Ser Ser Asn Pro Ala Ala Gly Leu His Leu Gln Leu Asn Tyr Thr Leu
    4485                4490                4495

CTG GAC GGC CAC TAC CTG TCT GAG GAA CCT GAG CCC TAC CTG GCA GTC     528
Leu Asp Gly His Tyr Leu Ser Glu Glu Pro Glu Pro Tyr Leu Ala Val
4500                4505                4510                4515

TAC CTA CAC TCG GAG CCC CGG CCC AAT GAG CAC AAC TGC TCG GCT AGC     576
Tyr Leu His Ser Glu Pro Arg Pro Asn Glu His Asn Cys Ser Ala Ser
                    4520                4525                4530

AGG AGG ATC CGC CCA GAG TCA CTC CAG GGT GCT GAC CAC CGG CCC TAC     624
Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly Ala Asp His Arg Pro Tyr
            4535                4540                4545

ACC TTC TTC ATT TCC CCG GGG AGC AGA GAC CCA GCG GGG AGT TAC CAT     672
Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp Pro Ala Gly Ser Tyr His
        4550                4555                4560

CTG AAC CTC TCC AGC CAC TTC CGC TGG TCG GCG CTG CAG GTG TCC GTG     720
Leu Asn Leu Ser Ser His Phe Arg Trp Ser Ala Leu Gln Val Ser Val
    4565                4570                4575
```

```
GGC CTG TAC ACG TCC CTG TGC CAG TAC TTC AGC GAG GAG GAC ATG GTG                768
Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe Ser Glu Glu Asp Met Val
4580                4585                4590                4595

TGG CGG ACA GAG GGG CTG CTG CCC CTG GAG GAG ACC TCG CCC CGC CAG               816
Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu Glu Thr Ser Pro Arg Gln
                    4600                4605                4610

GCC GTC TGC CTC ACC CGC CAC CTC ACC GCC TTC GGC GCC AGC CTC TTC               864
Ala Val Cys Leu Thr Arg His Leu Thr Ala Phe Gly Ala Ser Leu Phe
                4615                4620                4625

GTG CCC CCA AGC CAT GTC CGC TTT GTG TTT CCT GAG CCG ACA GCG GAT               912
Val Pro Pro Ser His Val Arg Phe Val Phe Pro Glu Pro Thr Ala Asp
        4630                4635                4640

GTA AAC TAC ATC GTC ATG CTG ACA TGT GCT GTG TGC CTG GTG ACC TAC               960
Val Asn Tyr Ile Val Met Leu Thr Cys Ala Val Cys Leu Val Thr Tyr
    4645                4650                4655

ATG GTC ATG GCC GCC ATC CTG CAC AAG CTG GAC CAG TTG GAT GCC AGC              1008
Met Val Met Ala Ala Ile Leu His Lys Leu Asp Gln Leu Asp Ala Ser
4660                4665                4670                4675

CGG GGC CGC GCC ATC CCT TTC TGT GGG CAG CGG GGC CGC TTC AAG TAC              1056
Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln Arg Gly Arg Phe Lys Tyr
                    4680                4685                4690

GAG ATC CTC GTC AAG ACA GGC TGG GGC CGG GGC TCA GGT ACC ACG GCC              1104
Glu Ile Leu Val Lys Thr Gly Trp Gly Arg Gly Ser Gly Thr Thr Ala
                4695                4700                4705

CAC GTG GGC ATC ATG CTG TAT GGG GTG GAC AGC CGG AGC GGC CAC CGG              1152
His Val Gly Ile Met Leu Tyr Gly Val Asp Ser Arg Ser Gly His Arg
        4710                4715                4720

CAC CTG GAC GGC GAC AGA GCC TTC CAC CGC AAC AGC CTG GAC ATC TTC              1200
His Leu Asp Gly Asp Arg Ala Phe His Arg Asn Ser Leu Asp Ile Phe
    4725                4730                4735

CGG ATC GCC ACC CCG CAC AGC CTG GGT AGC GTG TGG AAG ATC CGA GTG              1248
Arg Ile Ala Thr Pro His Ser Leu Gly Ser Val Trp Lys Ile Arg Val
4740                4745                4750                4755

TGG CAC GAC AAC AAA GGG CTC AGC CCT GCC TGG TTC CTG CAG CAC GTC              1296
Trp His Asp Asn Lys Gly Leu Ser Pro Ala Trp Phe Leu Gln His Val
                    4760                4765                4770

ATC GTC AGG GAC CTG CAG ACG GCA CGC AGC GCC TTC TTC GTG GTC AAT              1344
Ile Val Arg Asp Leu Gln Thr Ala Arg Ser Ala Phe Phe Leu Val Asn
                4775                4780                4785

GAC TGG CTT TCG GTG GAG ACG GAG GCC AAC GGG GGC CTG GTG GAG AAG              1392
Asp Trp Leu Ser Val Glu Thr Glu Ala Asn Gly Gly Leu Val Glu Lys
        4790                4795                4800

GAG GTG CTG GCC GCG AGC GAC GCA GCC CTT TTG CGC TTC CGG CGC CTG              1440
Glu Val Leu Ala Ala Ser Asp Ala Ala Leu Leu Arg Phe Arg Arg Leu
    4805                4810                4815

CTG GTG GCT GAG CTG CAG CGT GGC TTC TTT GAC AAG CAC ATC TGG CTC              1488
Leu Val Ala Glu Leu Gln Arg Gly Phe Phe Asp Lys His Ile Trp Leu
4820                4825                4830                4835
```

```
TCC ATA TGG GAC CGG CCG CCT CGT AGC CGT TTC ACT CGC ATC CAG AGG      1536
Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg Phe Thr Arg Ile Gln Arg
                4840                4845                4850

GCC ACC TGC TGC GTT CTC CTC ATC TGC CTC TTC CTG GGC GCC AAC GCC      1584
Ala Thr Cys Cys Val Leu Leu Ile Cys Leu Phe Leu Gly Ala Asn Ala
            4855                4860                4865

GTG TGG TAC GGG GCT GTT GGC GAC TCT GCC TAC AGC ACG GGG CAT GTG      1632
Val Trp Tyr Gly Ala Val Gly Asp Ser Ala Tyr Ser Thr Gly His Val
        4870                4875                4880

TCC AGG CTG AGC CCG CTG AGC GTC GAC ACA GTC GCT GTT GGC CTG GTG      1680
Ser Arg Leu Ser Pro Leu Ser Val Asp Thr Val Ala Val Gly Leu Val
    4885                4890                4895

TCC AGC GTG GTT GTC TAT CCC GTC TAC CTG GCC ATC CTT TTT CTC TTC      1728
Ser Ser Val Val Val Tyr Pro Val Tyr Leu Ala Ile Leu Phe Leu Phe
4900                4905                4910                4915

CGG ATG TCC CGG AGC AAG GTG GCT GGG AGC CCG AGC CCC ACA CCT GCC      1776
Arg Met Ser Arg Ser Lys Val Ala Gly Ser Pro Ser Pro Thr Pro Ala
                4920                4925                4930

GGG CAG CAG GTG CTG GAC ATC GAC AGC TGC CTG GAC TCG TCC GTG CTG      1824
Gly Gln Gln Val Leu Asp Ile Asp Ser Cys Leu Asp Ser Ser Val Leu
            4935                4940                4945

GAC AGC TCC TTC CTC ACG TTC TCA GGC CTC CAC GCT GAG GCC TTT GTT      1872
Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu His Ala Glu Ala Phe Val
        4950                4955                4960

GGA CAG ATG AAG AGT GAC TTG TTT CTG GAT GAT TCT AAG AGT CTG GTG      1920
Gly Gln Met Lys Ser Asp Leu Phe Leu Asp Asp Ser Lys Ser Leu Val
    4965                4970                4975

TGC TGG CCC TCC GGC GAG GGA ACG CTC AGT TGG CCG GAC CTG CTC AGT      1968
Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser Trp Pro Asp Leu Leu Ser
4980                4985                4990                4995

GAC CCG TCC ATT GTG GGT AGC AAT CTG CGG CAG CTG GCA CGG GGC CAG      2016
Asp Pro Ser Ile Val Gly Ser Asn Leu Arg Gln Leu Ala Arg Gly Gln
                5000                5005                5010

GCG GGC CAT GGG CTG GGC CCA GAG GAG GAC GGC TTC TCC CTG GCC AGC      2064
Ala Gly His Gly Leu Gly Pro Glu Glu Asp Gly Phe Ser Leu Ala Ser
            5015                5020                5025

CCC TAC TCG CCT GCC AAA TCC TTC TCA GCA TCA GAT GAA GAC CTG ATC      2112
Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala Ser Asp Glu Asp Leu Ile
        5030                5035                5040

CAG CAG GTC CTT GCC GAG GGG GTC AGC AGC CCA GCC CCT ACC CAA GAC      2160
Gln Gln Val Leu Ala Glu Gly Val Ser Ser Pro Ala Pro Thr Gln Asp
    5045                5050                5055

2ACC CAC ATG GAA ACG GAC CTG CTC AGC AGC CTG TCC AGC ACT CCT GGG      2208
Thr His Met Glu Thr Asp Leu Leu Ser Ser Leu Ser Ser Thr Pro Gly
5060                5065                5070                5075
```

```
GAG AAG ACA GAG ACG CTG GCG CTG CAG AGG CTG GGG GAG CTG GGG CCA      2256
Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg Leu Gly Glu Leu Gly Pro
                5080                5085                5090

8CCC AGC CCA GGC CTG AAC TGG GAA CAG CCC CAG GCA GCG AGG CTG TCC    2304
Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro Gln Ala Ala Arg Leu Ser
            5095                5100                5105

AGG ACA GGA CTG GTG GAG GGT CTG CGG AAG CGC CTG CTG CCG GCC TGG     2352
Arg Thr Gly Leu Val Glu Gly Leu Arg Lys Arg Leu Leu Pro Ala Trp
        5110                5115                5120

TGT GCC TCC CTG GCC CAC GGG CTC AGC CTG CTC CTG GTG GCT GTG GCT     2400
Cys Ala Ser Leu Ala His Gly Leu Ser Leu Leu Leu Val Ala Val Ala
    5125                5130                5135

GTG GCT GTC TCA GGG TGG GTG GGT GCG AGC TTC CCC CCG GGC GTG AGT     2448
Val Ala Val Ser Gly Trp Val Gly Ala Ser Phe Pro Pro Gly Val Ser
5140                5145                5150                5155

GTT GCG TGG CTC CTG TGC AGC AGC GCC AGC TTC CTG GCC TCA TTC CTC     2496
Val Ala Trp Leu Leu Ser Ser Ser Ala Ser Phe Leu Ala Ser Phe Leu
                5160                5165                5170

GGC TGG GAG CCA CTG AAG GTC TTG CTG GAA GCC CTG TAC TTC TCA CTG     2544
Gly Trp Glu Pro Leu Lys Val Leu Leu Glu Ala Leu Tyr Phe Ser Leu
            5175                5180                5185

GTG GCC AAG CGG CTG CAC CCG GAT GAA GAT GAC ACC CTG GTA GAG AGC     2592
Val Ala Lys Arg Leu His Pro Asp Glu Asp Asp Thr Leu Val Glu Ser
        5190                5195                5200

CCG GCT GTG ACG CCT GTG AGC GCA CGT GTG CCC CGC GTA CGG CCA CCC     2640
Pro Ala Val Thr Pro Val Ser Ala Arg Val Pro Arg Val Arg Pro Pro
    5205                5210                5215

CAC GGC TTT GCA CTC TTC CTG GCC AAG GAA GAA GCC CGC AAG GTC AAG     2688
His Gly Phe Ala Leu Phe Leu Ala Lys Glu Glu Ala Arg Lys Val Lys
5220                5225                5230                5235

AGG CTA CAT GGC ATG CTG CGG AGC CTC CTG GTG TAC ATG CTT TTT CTG     2736
Arg Leu His Gly Met Leu Arg Ser Leu Leu Val Tyr Met Leu Phe Leu
                5240                5245                5250

CTG GTG ACC CTG CTG GCC AGC TAT GGG GAT GCC TCA TGC CAT GGG CAC     2784
Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp Ala Ser Cys His Gly His
            5255                5260                5265

GCC TAC CGT CTG CAA AGC GCC ATC AAG CAG GAG CTG CAC AGC CGG GCC     2832
Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln Glu Leu His Ser Arg Ala
        5270                5275                5280

TTC CTG GCC ATC ACG CGG TCT GAG GAG CTC TGG CCA TGG ATG GCC CAC     2880
Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu Trp Pro Trp Met Ala His
    5285                5290                5295

GTG CTG CTG CCC TAC GTC CAC GGG AAC CAG TCC AGC CCA GAG CTG GGG     2928
Val Leu Leu Pro Tyr Val His Gly Asn Gln Ser Ser Pro Glu Leu Gly
5300                5305                5310                5315

CCC CCA CGG CTG CGG CAG GTG CGG CTG CAG GAA GCA CTC TAC CCA GAC     2976
```

```
CCC CCA CGG CTG CGG CAG GTG CGG CTG CAG GAA GCA CTC TAC CCA GAC      2976
Pro Pro Arg Leu Arg Gln Val Arg Leu Gln Glu Ala Leu Tyr Pro Asp
                5320                5325                5330

CCT CCC GGC CCC AGG GTC CAC ACG TGC TCG GCC GCA GGA GGC TTC AGC      3024
Pro Pro Gly Pro Arg Val His Thr Cys Ser Ala Ala Gly Gly Phe Ser
            5335                5340                5345

ACC AGC GAT TAC GAC GTT GGC TGG GAG AGT CCT CAC AAT GGC TCG GGG      3072
Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser Pro His Asn Gly Ser Gly
        5350                5355                5360

ACG TGG GCC TAT TCA GCG CCG GAT CTG CTG GGG GCA TGG TCC TGG GGC      3120
Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu Gly Ala Trp Ser Trp Gly
    5365                5370                5375

TCC TGT GCC GTG TAT GAC AGC GGG GGC TAC GTG CAG GAG CTG GGC CTG      3168
Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr Val Gln Glu Leu Gly Leu
5380                5385                5390                5395

AGC CTG GAG GAG AGC CGC GAC CGG CTG CGC TTC CTG CAG CTG CAC AAC      3216
Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg Phe Leu Gln Leu His Asn
                5400                5405                5410

TGG CTG GAC AAC AGG AGC CGC GCT GTG TTC CTG GAG CTC ACG CGC TAC      3264
Trp Leu Asp Asn Arg Ser Arg Ala Val Phe Leu Glu Leu Thr Arg Tyr
            5415                5420                5425

AGC CCG GCC GTG GGG CTG CAC GCC GCC GTC ACG CTG CGC CTC GAG TTC      3312
Ser Pro Ala Val Gly Leu His Ala Ala Val Thr Leu Arg Leu Glu Phe
        5430                5435                5440

CCG GCG GCC GGC CGC GCC CTG GCC GCC CTC AGC GTC CGC CCC TTT GCG      3360
Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu Ser Val Arg Pro Phe Ala
    5445                5450                5455

CTG CGC CGC CTC AGC GCG GGC CTC TCG CTG CCT CTG CTC ACC TCG GTG      3408
Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu Pro Leu Leu Thr Ser Val
5460                5465                5470                5475

TGC CTG CTG CTG TTC GCC GTG CAC TTC GCC GTG GCC GAG GCC CGT ACT      3456
Cys Leu Leu Leu Phe Ala Val His Phe Ala Val Ala Glu Ala Arg Thr
                5480                5485                5490

TGG CAC AGG GAA GGG CGC TGG CGC GTG CTG CGG CTC GGA GCC TGG GCG      3504
Trp His Arg Glu Gly Arg Trp Arg Val Leu Arg Leu Gly Ala Trp Ala
            5495                5500                5505

CGG TGG CTG CTG GTG GCG CTG ACG GCG GCC ACG GCA CTG GTA CGC CTC      3552
Arg Trp Leu Leu Val Ala Leu Thr Ala Ala Thr Ala Leu Val Arg Leu
        5510                5515                5520

GCC CAG CTG GGT GCC GCT GAC CGC CAG TGG ACC CGT TTC GTG CGC GGC      3600
Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp Thr Arg Phe Val Arg Gly
    5525                5530                5535

CGC CCG CGC CGC TTC ACT AGC TTC GAC CAG GTG GCG CAC GTG AGC TCC      3648
Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln Val Ala His Val Ser Ser
5540                5545                5550                5555
```

```
GCA GCC CGT GGC CTG GCG GCC TCG CTG CTC TTC CTG CTT TTG GTC AAG      3696
Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu Phe Leu Leu Leu Val Lys
                5560                5565                5570

2GCT GCC CAG CAC GTA CGC TTC GTG CGC CAG TGG TCC GTC TTT GGC AAG     3744
Ala Ala Gln His Val Arg Phe Val Arg Gln Trp Ser Val Phe Gly Lys
            5575                5580                5585

ACA TTA TGC CGA GCT CTG CCA GAG CTC CTG GGG GTC ACC TTG GGC CTG      3792
Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu Gly Val Thr Leu Gly Leu
        5590                5595                5600

GTG GTG CTC GGG GTA GCC TAC GCC CAG CTG GCC ATC CTG CTC GTG TCT      3840
Val Val Leu Gly Val Ala Tyr Ala Gln Leu Ala Ile Leu Leu Val Ser
    5605                5610                5615

TCC TGT GTG GAC TCC CTC TGG AGC GTG GCC CAG GCC CTG TTG GTG CTG      3888
Ser Cys Val Asp Ser Leu Trp Ser Val Ala Gln Ala Leu Leu Val Leu
5620                5625                5630                5635

TGC CCT GGG ACT GGG CTC TCT ACC CTG TGT CCT GCC GAG TCC TGG CAC      3936
Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys Pro Ala Glu Ser Trp His
                5640                5645                5650

CTG TCA CCC CTG CTG TGT GTG GGG CTC TGG GCA CTG CGG CTG TGG GGC      3984
Leu Ser Pro Leu Leu Cys Val Gly Leu Trp Ala Leu Arg Leu Trp Gly
            5655                5660                5665

GCC CTA CGG CTG GGG GCT GTT ATT CTC CGC TGG CGC TAC CAC GCC TTG      4032
Ala Leu Arg Leu Gly Ala Val Ile Leu Arg Trp Arg Tyr His Ala Leu
        5670                5675                5680

CGT GGA GAG CTG TAC CGG CCG GCC TGG GAG CCC CAG GAC TAC GAG ATG      4080
Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu Pro Gln Asp Tyr Glu Met
    5685                5690                5695

GTG GAG TTG TTC CTG CGC AGG CTG CGC CTC TGG ATG GGC CTC AGC AAG      4128
Val Glu Leu Phe Leu Arg Arg Leu Arg Leu Trp Met Gly Leu Ser Lys
5700                5705                5710                5715

GTC AAG GAG TTC CGC CAC AAA GTC CGC TTT GAA GGG ATG GAG CCG CTG      4176
Val Lys Glu Phe Arg His Lys Val Arg Phe Glu Gly Met Glu Pro Leu
                5720                5725                5730

CCC TCT CGC TCC TCC AGG GGC TCC AAG GTA TCC CCG GAT GTG CCC CCA      4224
Pro Ser Arg Ser Ser Arg Gly Ser Lys Val Ser Pro Asp Val Pro Pro
            5735                5740                5745

CCC AGC GCT GGC TCC GAT GCC TCG CAC CCC TCC ACC TCC TCC AGC CAG      4272
Pro Ser Ala Gly Ser Asp Ala Ser His Pro Ser Thr Ser Ser Ser Gln
        5750                5755                5760

CTG GAT GGG CTG AGC GTG AGC CTG GGC CGG CTG GGG ACA AGG TGT GAG      4320
Leu Asp Gly Leu Ser Val Ser Leu Gly Arg Leu Gly Thr Arg Cys Glu
    5765                5770                5775

CCT GAG CCC TCC CGC CTC CAA GCC GTG TTC GAG GCC CTG CTC ACC CAG      4368
Pro Glu Pro Ser Arg Leu Gln Ala Val Phe Glu Ala Leu Leu Thr Gln
5780                5785                5790                5795
```

```
TTT GAC CGA CTC AAC CAG GCC ACA GAG GAC GTC TAC CAG CTG GAG CAG      4416
Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp Val Tyr Gln Leu Glu Gln
                5800                5805                5810

CAG CTG CAC AGC CTG CAA GGC CGC AGG AGC AGC CGG GCG CCC GCC GGA      4464
Gln Leu His Ser Leu Gln Gly Arg Arg Ser Ser Arg Ala Pro Ala Gly
            5815                5820                5825

TCT TCC CGT GGC CCA TCC CCG GGC CTG CGG CCA GCA CTG CCC AGC CGC      4512
Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg Pro Ala Leu Pro Ser Arg
        5830                5835                5840

CTT GCC CGG GCC AGT CGG GGT GTG GAC CTG GCC ACT GGC CCC AGC AGG      4560
Leu Ala Arg Ala Ser Arg Gly Val Asp Leu Ala Thr Gly Pro Ser Arg
    5845                5850                5855

ACA CCT TCG GGC CAA GAA CAA GGT CCA CCC CAG CAG CAC TTA GTC CTC      4608
Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro Gln Gln His Leu Val Leu
5860                5865                5870                5875

CTT CCT GGC GGG GGT GGG CCG TGG AGT CGG AGT GGA CAC CGC TCA GTA      4656
Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg Ser Gly His Arg Ser Val
                5880                5885                5890

TTA CTT TCT GCC GCT GTC AAG GCC GAG GGC CAG GCA GAA TGG CTG CAC      4704
Leu Leu Ser Ala Ala Val Lys Ala Glu Gly Gln Ala Glu Trp Leu His
            5895                5900                5905

GTA GGT TCC CCA GAG AGC AGG CAG GGG CAT CTG TCT GTC TGT GGG CTT      4752
Val Gly Ser Pro Glu Ser Arg Gln Gly His Leu Ser Val Cys Gly Leu
        5910                5915                5920

CAG CAC TTT AAA GAG GCT GTG TGG CCA ACC AGG ACC CAG GGT CCC CTC      4800
Gln His Phe Lys Glu Ala Val Trp Pro Thr Arg Thr Gln Gly Pro Leu
    5925                5930                5935

CCC AGC TCC CTT GGG AAG GAC ACA GCA GTA TTG GAC GGT TTC              4842
Pro Ser Ser Leu Gly Lys Asp Thr Ala Val Leu Asp Gly Phe
5940                5945                5950

TAGCCTCTGA GATGCTAATT TATTTCCCCG AGTCCTCAGG TACAGCGGGC TGTGCCCGGC    4902

CCCACCCCCT GGGCAGATGT CCCCCACTGC TAAGGCTGCT GGCTTCAGGG AGGGTTAGCC    4962

2TGCACCGCCG CCACCCTGCC CCTAAGTTAT TACCTCTCCA GTTCCTACCG TACTCCCTGC    5022

ACCGTCTCAC TGTGTGTCTC GTGTCAGTAA TTTATATGGT GTTAAAATGT GTATATTTTT    5082

GTATGTCACT ATTTTCACTA GGGCTGAGGG GCCTGCGCCC AGAGCTGGCC TCCCCCAACA    5142

CCTGCTGCGC TTGGTAGGTG TGGTGGCGTT ATGGCAGCCC GGCTGCTGCT TGGATGCGAG    5202

CTTGGCCTTG GGCCGGTGCT GGGGGCACAG CTGTCTGCCA GGCACTCTCA TCACCCCAGA    5262

GGCCTTGTCA TCCTCCCTTG CCCCAGGCCA GGTAGCAAGA GAGCAGCGCC CAGGCCTGCT    5322

GGCATCAGGT CTGGGCAAGT AGCAGGACTA GGCATGTCAG AGGACCCCAG GGTGGTTAGA    5382

GGAAAAGACT CCTCCTGGGG GCTGGCTCCC AGGGTGGAGG AAGGTGACTG TGTGTGTGTG    5442

TGTGTGCGCG CGCGACGCGC GAGTGTGCTG TATGGCCCAG GCAGCCTCAA GGCCCTCGGA    5502
```

GCTGGCTGTG CCTGCTTCTG TGTACCACTT CTGTGGGCAT GGCCGCTTCT AGAGCCTCGA 5562

CACCCCCCCA ACCCCCGCAC CAAGCAGACA AAGTCAATAA AAGAGCTGTC TGACTGCAAA 5622

AAAAAAAAA 5631

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4: (Compare Figure 7)

Leu Asn Glu Glu Pro Leu Thr Leu Ala Gly Glu Glu Ile Val Ala Gln
1 5 10 15

Gly Lys Arg Ser Asp Pro Arg Ser Leu Leu Cys Tyr Gly Gly Ala Pro
20 25 30

Gly Pro Gly Cys His Phe Ser Ile Pro Glu Ala Phe Ser Gly Ala Leu
35 40 45

Ala Asn Leu Ser Asp Val Val Gln Leu Ile Phe Leu Val Asp Ser Asn
50 55 60

Pro Phe Pro Phe Gly Tyr Ile Ser Asn Tyr Thr Val Ser Thr Lys Val
65 70 75 80

Ala Ser Met Ala Phe Gln Thr Gln Ala Gly Ala Gln Ile Pro Ile Glu
85 90 95

Arg Leu Ala Ser Glu Arg Ala Ile Thr Val Lys Val Pro Asn Asn Ser
100 105 110

Asp Trp Ala Ala Arg Gly His Arg Ser Ser Ala Asn Ser Ala Asn Ser
115 120 125

Val Val Val Gln Pro Gln Ala Ser Val Gly Ala Val Val Thr Leu Asp
130 135 140

Ser Ser Asn Pro Ala Ala Gly Leu His Leu Gln Leu Asn Tyr Thr Leu
145 150 155 160

Leu Asp Gly His Tyr Leu Ser Glu Glu Pro Glu Pro Tyr Leu Ala Val
165 170 175

Tyr Leu His Ser Glu Pro Arg Pro Asn Glu His Asn Cys Ser Ala Ser
180 185 190

Arg Arg Ile Arg Pro Glu Ser Leu Gln Gly Ala Asp His Arg Pro Tyr
195 200 205

Thr Phe Phe Ile Ser Pro Gly Ser Arg Asp Pro Ala Gly Ser Tyr His
210 215 220

Leu Asn Leu Ser Ser His Phe Arg Trp Ser Ala Leu Gln Val Ser Val
225 230 235 240

Gly Leu Tyr Thr Ser Leu Cys Gln Tyr Phe Ser Glu Glu Asp Met Val
245 250 255

Trp Arg Thr Glu Gly Leu Leu Pro Leu Glu Glu Thr Ser Pro Arg Gln
260 265 270

Ala Val Cys Leu Thr Arg His Leu Thr Ala Phe Gly Ala Ser Leu Phe
275 280 285

Val Pro Pro Ser His Val Arg Phe Val Phe Pro Glu Pro Thr Ala Asp
290 295 300

Val Asn Tyr Ile Val Met Leu Thr Cys Ala Val Cys Leu Val Thr Tyr
305 310 315 320

Met Val Met Ala Ala Ile Leu His Lys Leu Asp Gln Leu Asp Ala Ser
325 330 335

Arg Gly Arg Ala Ile Pro Phe Cys Gly Gln Arg Gly Arg Phe Lys Tyr
340 345 350

Glu Ile Leu Val Lys Thr Gly Trp Gly Arg Gly Ser Gly Thr Thr Ala
355 360 365

His Val Gly Ile Met Leu Tyr Gly Val Asp Ser Arg Ser Gly His Arg
370 375 380

His Leu Asp Gly Asp Arg Ala Phe His Arg Asn Ser Leu Asp Ile Phe
385 390 395 400

Arg Ile Ala Thr Pro His Ser Leu Gly Ser Val Trp Lys Ile Arg Val
405 410 415

Trp His Asp Asn Lys Gly Leu Ser Pro Ala Trp Phe Leu Gln His Val
420 425 430

Ile Val Arg Asp Leu Gln Thr Ala Arg Ser Ala Phe Phe Leu Val Asn
435 440 445

Asp Trp Leu Ser Val Glu Thr Glu Ala Asn Gly Gly Leu Val Glu Lys
450 455 460

Glu Val Leu Ala Ala Ser Asp Ala Ala Leu Leu Arg Phe Arg Arg Leu
465 470 475 480

Leu Val Ala Glu Leu Gln Arg Gly Phe Phe Asp Lys His Ile Trp Leu
485 490 495

Ser Ile Trp Asp Arg Pro Pro Arg Ser Arg Phe Thr Arg Ile Gln Arg
500 505 510

Ala Thr Cys Cys Val Leu Leu Ile Cys Leu Phe Leu Gly Ala Asn Ala
515 520 525

Val Trp Tyr Gly Ala Val Gly Asp Ser Ala Tyr Ser Thr Gly His Val
530 535 540

Ser Arg Leu Ser Pro Leu Ser Val Asp Thr Val Ala Val Gly Leu Val
545 550 555 560

Ser Ser Val Val Val Tyr Pro Val Tyr Leu Ala Ile Leu Phe Leu Phe
565 570 575

Arg Met Ser Arg Ser Lys Val Ala Gly Ser Pro Ser Pro Thr Pro Ala
580 585 590

Gly Gln Gln Val Leu Asp Ile Asp Ser Cys Leu Asp Ser Ser Val Leu
595 600 605

Asp Ser Ser Phe Leu Thr Phe Ser Gly Leu His Ala Glu Ala Phe Val
610 615 620

Gly Gln Met Lys Ser Asp Leu Phe Leu Asp Asp Ser Lys Ser Leu Val
625 630 635 640

Cys Trp Pro Ser Gly Glu Gly Thr Leu Ser Trp Pro Asp Leu Leu Ser
645 650 655

Asp Pro Ser Ile Val Gly Ser Asn Leu Arg Gln Leu Ala Arg Gly Gln
660 665 670

Ala Gly His Gly Leu Gly Pro Glu Glu Asp Gly Phe Ser Leu Ala Ser
675 680 685

Pro Tyr Ser Pro Ala Lys Ser Phe Ser Ala Ser Asp Glu Asp Leu Ile
690 695 700

Gln Gln Val Leu Ala Glu Gly Val Ser Ser Pro Ala Pro Thr Gln Asp
705 710 715 720

Thr His Met Glu Thr Asp Leu Leu Ser Ser Leu Ser Ser Thr Pro Gly
725 730 735

Glu Lys Thr Glu Thr Leu Ala Leu Gln Arg Leu Gly Glu Leu Gly Pro
740 745 750

Pro Ser Pro Gly Leu Asn Trp Glu Gln Pro Gln Ala Ala Arg Leu Ser
755 760 765

Arg Thr Gly Leu Val Glu Gly Leu Arg Lys Arg Leu Leu Pro Ala Trp
770 775 780

Cys Ala Ser Leu Ala His Gly Leu Ser Leu Leu Leu Val Ala Val Ala
785 790 795 800

Val Ala Val Ser Gly Trp Val Gly Ala Ser Phe Pro Pro Gly Val Ser
805 810 815

Val Ala Trp Leu Leu Ser Ser Ser Ala Ser Phe Leu Ala Ser Phe Leu
820 825 830

Gly Trp Glu Pro Leu Lys Val Leu Leu Glu Ala Leu Tyr Phe Ser Leu
835 840 845

Val Ala Lys Arg Leu His Pro Asp Glu Asp Asp Thr Leu Val Glu Ser
850 855 860

Pro Ala Val Thr Pro Val Ser Ala Arg Val Pro Arg Val Arg Pro Pro
865 870 875 880

His Gly Phe Ala Leu Phe Leu Ala Lys Glu Glu Ala Arg Lys Val Lys
885 890 895

Arg Leu His Gly Met Leu Arg Ser Leu Leu Val Tyr Met Leu Phe Leu
900 905 910

Leu Val Thr Leu Leu Ala Ser Tyr Gly Asp Ala Ser Cys His Gly His
915 920 925

Ala Tyr Arg Leu Gln Ser Ala Ile Lys Gln Glu Leu His Ser Arg Ala
930 935 940

Phe Leu Ala Ile Thr Arg Ser Glu Glu Leu Trp Pro Trp Met Ala His
945 950 955 960

Val Leu Leu Pro Tyr Val His Gly Asn Gln Ser Ser Pro Glu Leu Gly
965 970 975

Pro Pro Arg Leu Arg Gln Val Arg Leu Gln Glu Ala Leu Tyr Pro Asp
980 985 990

Pro Pro Gly Pro Arg Val His Thr Cys Ser Ala Ala Gly Gly Phe Ser
995 1000 1005

Thr Ser Asp Tyr Asp Val Gly Trp Glu Ser Pro His Asn Gly Ser Gly
1010 1015 1020

Thr Trp Ala Tyr Ser Ala Pro Asp Leu Leu Gly Ala Trp Ser Trp Gly
1025 1030 1035 1040

Ser Cys Ala Val Tyr Asp Ser Gly Gly Tyr Val Gln Glu Leu Gly Leu
1045 1050 1055

2
Ser Leu Glu Glu Ser Arg Asp Arg Leu Arg Phe Leu Gln Leu His Asn
1060 1065 1070

Trp Leu Asp Asn Arg Ser Arg Ala Val Phe Leu Glu Leu Thr Arg Tyr
1075 1080 1085

Ser Pro Ala Val Gly Leu His Ala Ala Val Thr Leu Arg Leu Glu Phe
1090 1095 1100

Pro Ala Ala Gly Arg Ala Leu Ala Ala Leu Ser Val Arg Pro Phe Ala
1105 1110 1115 1120

Leu Arg Arg Leu Ser Ala Gly Leu Ser Leu Pro Leu Leu Thr Ser Val
1125 1130 1135

Cys Leu Leu Leu Phe Ala Val His Phe Ala Val Ala Glu Ala Arg Thr
1140 1145 1150

Trp His Arg Glu Gly Arg Trp Arg Val Leu Arg Leu Gly Ala Trp Ala
1155 1160 1165

Arg Trp Leu Leu Val Ala Leu Thr Ala Ala Thr Ala Leu Val Arg Leu
1170 1175 1180

8
Ala Gln Leu Gly Ala Ala Asp Arg Gln Trp Thr Arg Phe Val Arg Gly
1185 1190 1195 1200

Arg Pro Arg Arg Phe Thr Ser Phe Asp Gln Val Ala His Val Ser Ser
2 1205 1210 1215

Ala Ala Arg Gly Leu Ala Ala Ser Leu Leu Phe Leu Leu Leu Val Lys
1220 1225 1230

Ala Ala Gln His Val Arg Phe Val Arg Gln Trp Ser Val Phe Gly Lys
1235 1240 1245

Thr Leu Cys Arg Ala Leu Pro Glu Leu Leu Gly Val Thr Leu Gly Leu
1250 1255 1260

Val Val Leu Gly Val Ala Tyr Ala Gln Leu Ala Ile Leu Leu Val Ser
1265 1270 1275 1280

Ser Cys Val Asp Ser Leu Trp Ser Val Ala Gln Ala Leu Leu Val Leu
1285 1290 1295

Cys Pro Gly Thr Gly Leu Ser Thr Leu Cys Pro Ala Glu Ser Trp His
1300 1305 1310

Leu Ser Pro Leu Leu Cys Val Gly Leu Trp Ala Leu Arg Leu Trp Gly
1315 1320 1325

Ala Leu Arg Leu Gly Ala Val Ile Leu Arg Trp Arg Tyr His Ala Leu
1330 1335 1340

Arg Gly Glu Leu Tyr Arg Pro Ala Trp Glu Pro Gln Asp Tyr Glu Met
1345 1350 1355 1360

Val Glu Leu Phe Leu Arg Arg Leu Arg Leu Trp Met Gly Leu Ser Lys
1365 1370 1375

Val Lys Glu Phe Arg His Lys Val Arg Phe Glu Gly Met Glu Pro Leu
1380 1385 1390

Pro Ser Arg Ser Ser Arg Gly Ser Lys Val Ser Pro Asp Val Pro Pro
1395 1400 1405

Pro Ser Ala Gly Ser Asp Ala Ser His Pro Ser Thr Ser Ser Ser Gln
1410 1415 1420

Leu Asp Gly Leu Ser Val Ser Leu Gly Arg Leu Gly Thr Arg Cys Glu
1425 1430 1435 1440

Pro Glu Pro Ser Arg Leu Gln Ala Val Phe Glu Ala Leu Leu Thr Gln
1445 1450 1455

Phe Asp Arg Leu Asn Gln Ala Thr Glu Asp Val Tyr Gln Leu Glu Gln
1460 1465 1470

Gln Leu His Ser Leu Gln Gly Arg Arg Ser Ser Arg Ala Pro Ala Gly
1475 1480 1485

Ser Ser Arg Gly Pro Ser Pro Gly Leu Arg Pro Ala Leu Pro Ser Arg
1490 1495 1500

Leu Ala Arg Ala Ser Arg Gly Val Asp Leu Ala Thr Gly Pro Ser Arg
1505 1510 1515 1520

Thr Pro Ser Gly Gln Glu Gln Gly Pro Pro Gln Gln His Leu Val Leu
1525 1530 1535

Leu Pro Gly Gly Gly Gly Pro Trp Ser Arg Ser Gly His Arg Ser Val
1540 1545 1550

Leu Leu Ser Ala Ala Val Lys Ala Glu Gly Gln Ala Glu Trp Leu His
1555 1560 1565

Val Gly Ser Pro Glu Ser Arg Gln Gly His Leu Ser Val Cys Gly Leu
1570 1575 1580

Gln His Phe Lys Glu Ala Val Trp Pro Thr Arg Thr Gln Gly Pro Leu
1585 1590 1595 1600

Pro Ser Ser Leu Gly Lys Asp Thr Ala Val Leu Asp Gly Phe
1605 1610

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5: (Compare Figure 8)

AGCTTGGCAC CATCAAGGGC CAGTTCAACT TTGTCCACGT GATCGTCACC CCGCTGGACT 60

ACGAGTGCAA CCTGGTGTCC CTGCAGTGCA GGAAAGACAT GGAGGGCCTT GTGGACACCA 120

GCGTGGCCAA GATCGTGTCT GACCGCAACC TGCCCTTCGT GGCCCGCCAG ATGGCCCTGC 180

ACGCAAATAT GGCCTCACAG GTGCATCATA GCCGCTCCAA CCCCACCGAT ATCTACCCCT 240

CCAAGTGGAT TGCCCGGCTC CGCCACATCA AGCGGCTCCG CCAGCGGATC TGCGAGGAAG 300

CCGCCTACTC CAACCCCAGC CTACCTCTGG TGCACCCTCC GTCCCATAGC AAAGCCCCTG 360

CACAGACTCC AGCCGAGCCC ACACCTGGCT ATGAGGTGGG CCAGCGGAAG CGCCTCATCT 420

CCTCGGTGGA GGACTTCACC GAGTTTGTGT GAGGCCGGGG CCCTCCCTCC TGCACTGGCC 480

TTGGACGGTA TTGCCTGTCA GTGAAATAAA TAAAGTCCTG ACCCCAGTGC ACAGACATAG 540

AGGCACAGAT TGC 553

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6: (Compare Figure 9)

CTGGTGTGTG TGAGACGTGC GGGGCTGGGA AGTGTTGGCA GAGCCGCGAG TACCGTCCTC 60

ACTCCTTTTG TTCTTTTGAC GTAAGCTGGC GAGTGGCACT GCCTGAGTTC CGCTCAGTGC 120

CCGCCCTGAT GTGCGGACCC CGCTGCATTC TTGCTGTTAG GTGGTGGCGG TGTGCGCTGT 180

CGCTGGTGGG CACCGAGAGT CTTTGGGAGC TTTGGGGAGG TTGTGCCAAG CCTGAGCCTC 240

GACGTCCCCC TTCCCGGCTT TCTGTTGGCT CTTCTGAGGC CAGGGCATCT CTATGAGGGC 300

CTCCTGCTGG AGCCGTCTCT GTGGATCTCC TCTGCCATCC TGGCCCATGA GTGGGTGATG 360

CGCTGGCCAC CATCTGGTGA CAGTGGCCGG GCACCGCTGC CAAATGTGGG TCCCGCATCT 420

GCAAGCCCCT CCCTGGGTCC CCTAGGGTAT GGGGTGGTTC TGCCACTGCC CTCGCTCCCC 480

CACCTTGGGG TGCCTCTCCC CCTGCTCGTG GGGGAGA 517

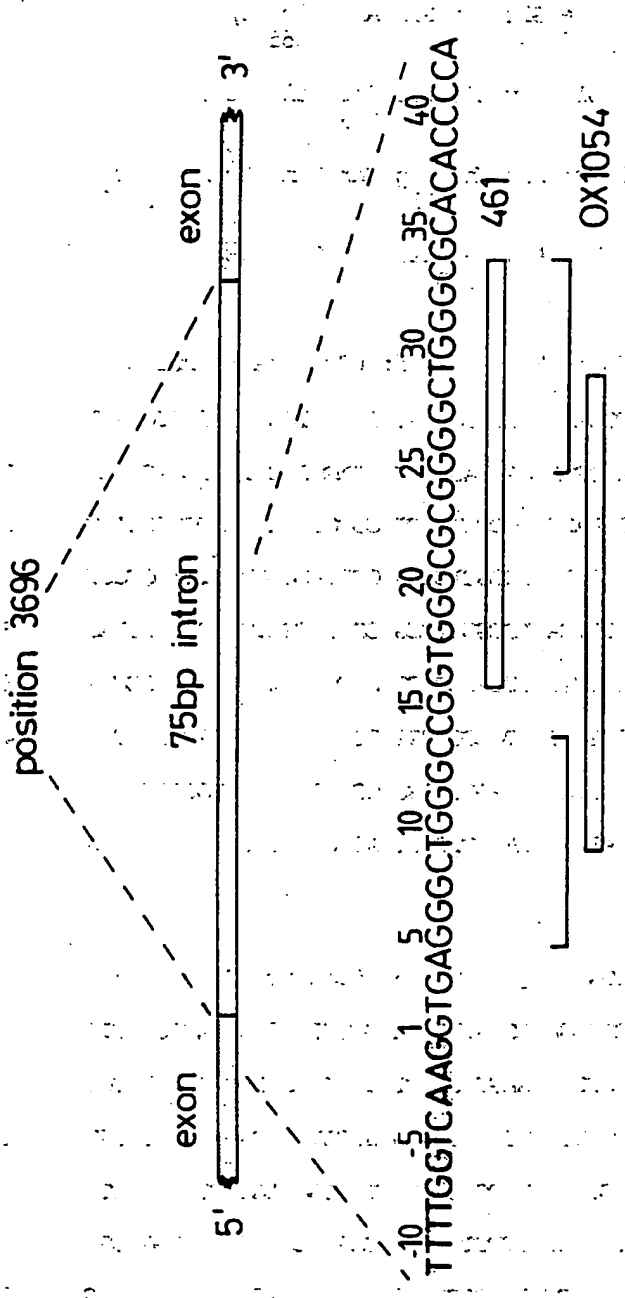
position 3696
exon
75bp intron
exon
5'
3'
-10 TTTTGGTCAAGGTGAGGGCTGGGCCGGTGGGCGCGGGGCTGGGCGCACACCCCA 40
-5
1
5
10
15
20
25
30
35
461
OX1054

Fig. 11

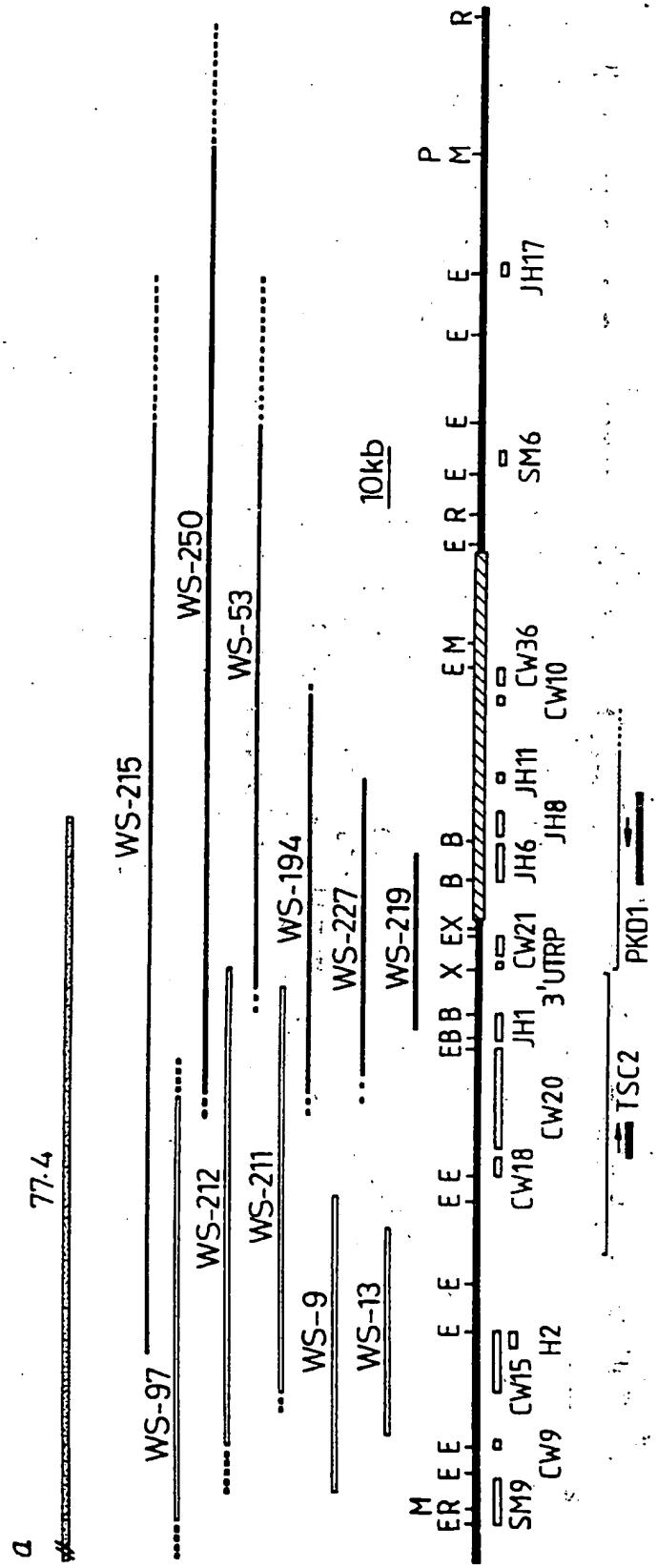
a
77.4
WS-97
WS-215
WS-212
WS-250
WS-211
WS-53
WS-9
WS-194
WS-13
WS-227
WS-219
10kb
M
ER E E
SM9
CW9
E E
CW15
H2
E E
CW18
JH1
EB B X EX B B
CW21
3'UTRP
CW20
JH6
JH8
JH11
EM
CW36
CW10
E R E E E E
SM6
JH17
P
M
R
TSC2
PKD1

Fig. 12

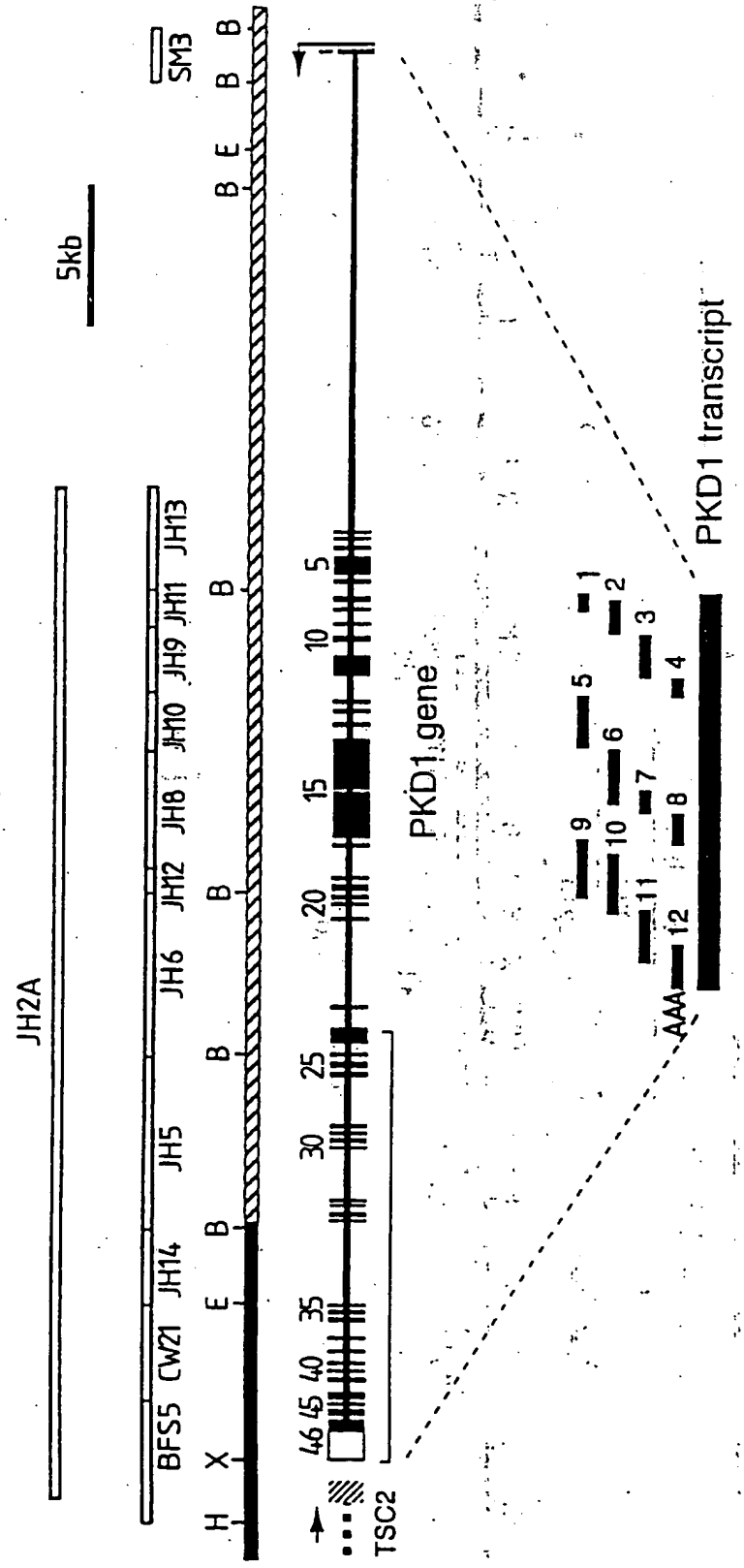
5kb
SM3
B E B B
JH2A
JH13 JH11 JH9 JH10 JH8 JH12 JH6 JH5 JH14 CW21 BFS5
H X E B B B B
1 5 10 15 20 25 30 35 40 45 46
TSC2
PKD1 gene
1 2 3 4 5 6 7 8 9 10 11 12 AAA
PKD1 transcript

Fig. 13

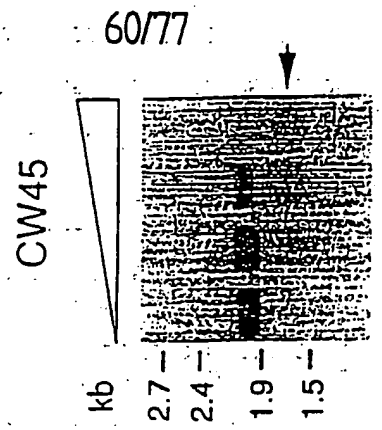

Fig.14b

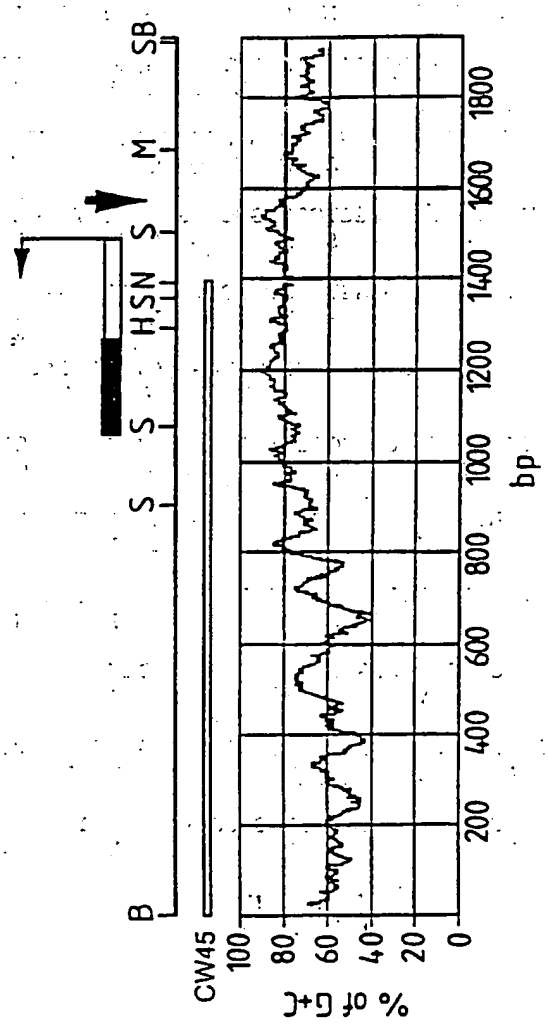

Fig.14a

```
   1  GCACTGCAGCGCCAGCGTCCGAGCGGGCGGCCGAGCTCCCGGAGCGGCCTGGCCCCGAGC   60

  61  CCCGAGCGGGCGTCGCTCAGCAGCAGGTCGCGGCCGCGCAGCCCCATCCAGCCCCGCGCC  120

 121  CGCCATGCCGTCCGCGGGCCCCGCCTGAGCTGCGGTCTCCGCGCGCGGGCGGGCCTGGGG  180

 181  ACGGCGGGGCCATGCGCGCGCTGCCCTAACGATGCCGCCCGCCGCGCCCGCCCGCCTGGC  240
   1                                M  P  P  A  A  P  A  R  L  A      10

 241  GCTGGCCCTGGGCCTGGGCCTGTGGCTCGGGGGCGCTGGCGGGGGGCCCCGGGCGCGGCTG  300
  11   L  A  L  G  L  G  L  W  L  G  A  L  A  G  G  P  G  R  G  C   30
                                          ▲

 301  CGGGCCCTGCGAGCCCCCCTGCCTCTGCGGCCCAGCGCCCGGCGCCGCCTGCCGCGTCAA  360
  31   G  P  C  E  P  P  C  L  C  G  P  A  P  G  A  A  C  R  V  N   50
                                                                  *

 361  CTGCTCGGGCCGCGGGCTGCGGACGCTCGGTCCCGCGCTGCGCATCCCCGCGGACGCCAC  420
  51   C  S  G  R  G  L  R  T  L  G  P  A  L  R  I  P  A  D  A  T   70

 421  AGCGCTAGACGTCTCCCACAACCTGCTCCGGGCGCTGGACGTTGGGCTCCTGGCGAACCT  480
  71   A  L  D  V  S  H  N  L  L  R  A  L  D  V  G  L  L  A  N  L   90
                                                                  *

 481  CTCGGCGCTGGCAGAGCTGGATATAAGCAACAACAAGATTTCTACGTTAGAAGAAGGAAT  540
  91   S  A  L  A  E  L  D  I  S  N  N  K  I  S  T  L  E  E  G  I  110

 541  ATTTGCTAATTTATTTAATTTAAGTGAAATAAACCTGAGTGGGAACCCGTTTGAGTGTGA  600
 111   F  A  N  L  F  N  L  S  E  I  N  L  S  G  N  P  F  E  C  D  130
                     *              *

 601  CTGTGGCCTGGCGTGGCTGCCGCGATGGGCGGAGGAGCAGCAGGTGCGGGTGGTGCAGCC  660
 131   C  G  L  A  W  L  P  R  W  A  E  E  Q  Q  V  R  V  V  Q  P  150

 661  CGAGGCAGCCACGTGTGCTGGGCCTGGCTCCCTGGCTGGCCAGCCTCTGCTTGGCATCCC  720
 151   E  A  A  T  C  A  G  P  G  S  L  A  G  Q  P  L  L  G  I  P  170

 721  CTTGCTGGACAGTGGCTGTGGTGAGGAGTATGTCGCCTGCCTCCCTGACAACAGCTCAGG  780
 171   L  L  D  S  G  C  G  E  E  Y  V  A  C  L  P  D  N  S  S  G  190
                                                    *

 781  CACCGTGGCAGCAGTGTCCTTTTCAGCTGCCCACGAAGGCCTGCTTCAGCCAGAGGCCTG  840
 191   T  V  A  A  V  S  F  S  A  A  H  E  G  L  L  Q  P  E  A  C  210

 841  CAGCGCCTTCTGCTTCTCCACCGGCCAGGGCCTCGCAGCCCTCTCGGAGCAGGGCTGGTG  900
 211   S  A  F  C  F  S  T  G  Q  G  L  A  A  L  S  E  Q  G  W  C  230

 901  CCTGTGTGGGGCGGCCCAGCCCTCCAGTGCCTCCTTTGCCTGCCTGTCCCTCTGCTCCGG  960
 231   L  C  G  A  A  Q  P  S  S  A  S  F  A  C  L  S  L  C  S  G  250

 961  CCCCCCGCCACCTCCTGCCCCCACCTGTAGGGGCCCCACCCTCCTCCAGCACGTCTTCCC 1020
 251   P  P  P  P  P  A  P  T  C  R  G  P  T  L  L  Q  H  V  F  P  270

1021  TGCCTCCCCAGGGGCCACCCTGGTGGGGCCCCACGGACCTCTGGCCTCTGGCCAGCTAGC 1080
 271   A  S  P  G  A  T  L  V  G  P  H  G  P  L  A  S  G  Q  L  A  290

1081  AGCCTTCCACATCGCTGCCCCGCTCCCTGTCACTGCCACACGCTGGGACTTCGGAGACGG 1140
 291   A  F  H  I  A  A  P  L  P  V  T  A  T  R  W  D  F  G  D  G  310

1141  CTCCGCCGAGGTGGATGCCGCTGGGCCGGCTGCCTCGCATCGCTATGTGCTGCCTGGGCG 1200
 311   S  A  E  V  D  A  A  G  P  A  A  S  H  R  Y  V  L  P  G  R  330
```

Fig. 15

```
1201  CTATCACGTGACGGCCGTGCTGGCCCTGGGGGCCGGCTCAGCCCTGCTGGGGACAGACGT  1260
 331   Y  H  V  T  A  V  L  A  L  G  A  G  S  A  L  L  G  T  D  V   350

1261  GCAGGTGGAAGCGGCACCTGCCGCCCTGGAGCTCGTGTGCCCGTCCTCGGTGCAGAGTGA  1320
 351   Q  V  E  A  A  P  A  A  L  E  L  V  C  P  S  S  V  Q  S  D   370

1321  CGAGAGCCTTGACCTCAGCATCCAGAACCGCGGTGGTTCAGGCCTGGAGGCCGCCTACAG  1380
 371   E  S  L  D  L  S  I  Q  N  R  G  G  S  G  L  E  A  A  Y  S   390

1381  CATCGTGGCCCTGGGCGAGGAGCCGGCCCGAGCGGTGCACCCGCTCTGCCCCTCGGACAC  1440
 391   I  V  A  L  G  E  E  P  A  R  A  V  H  P  L  C  P  S  D  T   410

1441  GGAGATCTTCCCTGGCAACGGGCACTGCTACCGCCTGGTGGTGGAGAAGGCGGCCTGGCT  1500
 411   E  I  F  P  G  N  G  H  C  Y  R  L  V  V  E  K  A  A  W  L   430

1501  GCAGGCGCAGGAGCAGTGTCAGGCCTGGGCCGGGGCCGCCCTGGCAATGGTGGACAGTCC  1560
 431   Q  A  Q  E  Q  C  Q  A  W  A  G  A  A  L  A  M  V  D  S  P   450

1561  CGCCGTGCAGCGCTTCCTGGTCTCCCGGGTCACCAGGAGCCTAGACGTGTGGATCGGCTT  1620
 451   A  V  Q  R  F  L  V  S  R  V  T  R  S  L  D  V  W  I  G  F   470

1621  CTCGACTGTGCAGGGGGTGGAGGTGGGCCCAGCGCCGCAGGGCGAGGCCTTCAGCCTGGA  1680
 471   S  T  V  Q  G  V  E  V  G  P  A  P  Q  G  E  A  F  S  L  E   490

1681  GAGCTGCCAGAACTGGCTGCCCGGGGAGCCACACCCAGCCACAGCCGAGCACTGCGTCCG  1740
 491   S  C  Q  N  W  L  P  G  E  P  H  P  A  T  A  E  H  C  V  R   510

1741  GCTCGGGCCCACCGGGTGGTGTAACACCGACCTGTGCTCAGCGCCGCACAGCTACGTCTG  1800
 511   L  G  P  T  G  W  C  N  T  D  L  C  S  A  P  H  S  Y  V  C   530

1801  CGAGCTGCAGCCCGGAGGCCCAGTGCAGGATGCCGAGAACCTCCTCGTGGGAGCGCCCAG  1860
 531   E  L  Q  P  G  G  P  V  Q  D  A  E  N  L  L  V  G  A  P  S   550

1861  TGGGGACCTGCAGGGACCCCTGACGCCTCTGGCACAGCAGGACGGCCTCTCAGCCCCGCA  1920
 551   G  D  L  Q  G  P  L  T  P  L  A  Q  Q  D  G  L  S  A  P  H   570

1921  CGAGCCCGTGGAGGTCATGGTATTCCCGGGCCTGCGTCTGAGCCGTGAAGCCTTCCTCAC  1980
 571   E  P  V  E  V  M  V  F  P  G  L  R  L  S  R  E  A  F  L  T   590

1981  CACGGCCGAATTTGGGACCCAGGAGCTCCGGCGGCCCGCCCAGCTGCGGCTGCAGGTGTA  2040
 591   T  A  E  F  G  T  Q  E  L  R  R  P  A  Q  L  R  L  Q  V  Y   610

2041  CCGGCTCCTCAGCACAGCAGGGACCCCGGAGAACGGCAGCGAGCCTGAGAGCAGGTCCCC  2100
 611   R  L  L  S  T  A  G  T  P  E  N  G  S  E  P  E  S  R  S  P   630
                                      *

2101  GGACAACAGGACCCAGCTGGCCCCCGCGTGCATGCCAGGGGGACGCTGGTGCCCTGGAGC  2160
 631   D  N  R  T  Q  L  A  P  A  C  M  P  G  G  R  W  C  P  G  A   650
          *

2161  CAACATCTGCTTGCCGCTGGACGCCTCTTGCCACCCCCAGGCCTGCGCCAATGGCTGCAC  2220
 651   N  I  C  L  P  L  D  A  S  C  H  P  Q  A  C  A  N  G  C  T   670

2221  GTCAGGGCCAGGGCTACCCGGGGCCCCCTATGCGCTATGGAGAGAGTTCCTCTTCTCCGT  2280
 671   S  G  P  G  L  P  G  A  P  Y  A  L  W  R  E  F  L  F  S  V   690

2281  TGCCGCGGGGCCCCCCGCGCAGTACTCGGTCACCCTCCACGGCCAGGATGTCCTCATGCT  2340
 691   A  A  G  P  P  A  Q  Y  S  V  T  L  H  G  Q  D  V  L  M  L   710

2341  CCCTGGTGACCTCGTTGGCTTGCAGCACGACGCTGGCCCTGGCGCCCTCCTGCACTGCTC  2400
 711   P  G  D  L  V  G  L  Q  H  D  A  G  P  G  A  L  L  H  C  S   730
```

```
2401  GCCGGCTCCCGGCCACCCTGGTCCCCAGGCCCCGTACCTCTCCGCCAACGCCTCGTCATG  2460
 731   P  A  P  G  H  P  G  P  Q  A  P  Y  L  S  A  N  A  S  S  W   750
                                                    *
2461  GCTGCCCCACTTGCCAGCCCAGCTGGAGGGCACTTGGGCCTGCCCTGCCTGTGCCCTGCG  2520
 751   L  P  H  L  P  A  Q  L  E  G  T  W  A  C  P  A  C  A  L  R   770

2521  GCTGCTTGCAGCCACGGAACAGCTCACCGTGCTGCTGGGCTTGAGGCCCAACCCTGGACT  2580
 771   L  L  A  A  T  E  Q  L  T  V  L  L  G  L  R  P  N  P  G  L   790

2581  GCGGATGCCTGGGCGCTATGAGGTCCGGGCAGAGGTGGGCAATGGCGTGTCCAGGCACAA  2640
 791   R  M  P  G  R  Y  E  V  R  A  E  V  G  N  G  V  S  R  H  N   810
                                                                *
2641  CCTCTCCTGCAGCTTTGACGTGGTCTCCCCAGTGGCTGGGCTGCGGGTCATCTACCCTGC  2700
 811   L  S  C  S  F  D  V  V  S  P  V  A  G  L  R  V  I  Y  P  A   830

2701  CCCCCGCGACGGCCGCCTCTACGTGCCCACCAACGGCTCAGCCTTGGTGCTCCAGGTGGA  2760
 831   P  R  D  G  R  L  Y  V  P  T  N  G  S  A  L  V  L  Q  V  D   850
                                       *
2761  CTCTGGTGCCAACGCCACGGCCACGGCTCGCTGGCCTGGGGGCAGTGTCAGCGCCCGCTT  2820
 851   S  G  A  N  A  T  A  T  A  R  W  P  G  G  S  V  S  A  R  F   870
             *
2821  TGAGAATGTCTGCCCTGCCCTGGTGGCCACCTTCGTGCCCGGCTGCCCCTGGGAGACCAA  2880
 871   E  N  V  C  P  A  L  V  A  T  F  V  P  G  C  P  W  E  T  N   890
                                                                *
2881  CGATACCCTGTTCTCAGTGGTAGCACTGCCGTGGCTCAGTGAGGGGGAGCACGTGGTGGA  2940
 891   D  T  L  F  S  V  V  A  L  P  W  L  S  E  G  E  H  V  V  D   910

2941  CGTGGTGGTGGAAAACAGCGCCAGCCGGGCCAACCTCAGCCTGCGGGTGACGGCGGAGGA  3000
 911   V  V  V  E  N  S  A  S  R  A  N  L  S  L  R  V  T  A  E  E   930
                                       *
3001  GCCCATCTGTGGCCTCCGCGCCACGCCCAGCCCCGAGGCCCGTGTACTGCAGGGAGTCCT  3060
 931   P  I  C  G  L  R  A  T  P  S  P  E  A  R  V  L  Q  G  V  L   950

3061  AGTGAGGTACAGCCCCGTGGTGGAGGCCGGCTCGGACATGGTCTTCCGGTGGACCATCAA  3120
 951   V  R  Y  S  P  V  V  E  A  G  S  D  M  V  F  R  W  T  I  N   970

3121  CGACAAGCAGTCCCTGACCTTCCAGAACGTGGTCTTCAATGTCATTTATCAGAGCGCGGC  3180
 971   D  K  Q  S  L  T  F  Q  N  V  V  F  N  V  I  Y  Q  S  A  A   990

3181  GGTCTTCAAGCTCTCACTGACGGCCTCCAACCACGTGAGCAACGTCACCGTGAACTACAA  3240
 991   V  F  K  L  S  L  T  A  S  N  H  V  S  N  V  T  V  N  Y  N  1010
                                                    *
3241  CGTAACCGTGGAGCGGATGAACAGGATGCAGGGTCTGCAGGTCTCCACAGTGCCGGCCGT  3300
1011   V  T  V  E  R  M  N  R  M  Q  G  L  Q  V  S  T  V  P  A  V  1030

3301  GCTGTCCCCCAATGCCACACTGGTACTGACGGGTGGTGTGCTGGTGGACTCAGCTGTGGA  3360
1031   L  S  P  N  A  T  L  V  L  T  G  G  V  L  V  D  S  A  V  E  1050
             *
3361  GGTGGCCTTCCTGTGGAACTTTGGGGATGGGGAGCAGGCCCTCCACCAGTTCCAGCCTCC  3420
1051   V  A  F  L  W  N  F  G  D  G  E  Q  A  L  H  Q  F  Q  P  P  1070

3421  GTACAACGAGTCCTTCCCGGTTCCAGACCCCTCGGTGGCCCAGGTGCTGGTGGAGCACAA  3480
1071   Y  N  E  S  F  P  V  P  D  P  S  V  A  Q  V  L  V  E  H  N  1090
          *
3481  TGTCATGCACACCTACGCTGCCCCAGGTGAGTACCTCCTGACCGTGCTGGCATCTAATGC  3540
1091   V  M  H  T  Y  A  A  P  G  E  Y  L  L  T  V  L  A  S  N  A  1110

3541  CTTCGAGAACCTGACGCAGCAGGTGCCTGTGAGCGTGCGCGCCTCCCTGCCCTCCGTGGC  3600
1111   F  E  N  L  T  Q  Q  V  P  V  S  V  R  A  S  L  P  S  V  A  1130
             *
```

```
3601  TGTGGGTGTGAGTGACGGCGTCCTGGTGGCCGGCCGGCCCGTCACCTTCTACCCGCACCC  3660
1131   V  G  V  S  D  G  V  L  V  A  G  R  P  V  T  F  Y  P  H  P   1150

3661  GCTGCCCTCGCCTGGGGGTGTTCTTTACACGTGGGACTTCGGGGACGGCTCCCCTGTCCT  3720
1151   L  P  S  P  G  G  V  L  Y  T  W  D  F  G  D  G  S  P  V  L   1170

3721  GACCCAGAGCCAGCCGGCTGCCAACCACACCTATGCCTCGAGGGGCACCTACCACGTGCG  3780
1171   T  Q  S  Q  P  A  A  N  H  T  Y  A  S  R  G  T  Y  H  V  R   1190
                          *

3781  CCTGGAGGTCAACAACACGGTGAGCGGTGCGGCGGCCCAGGCGGATGTGCGCGTCTTTGA  3840
1191   L  E  V  N  N  T  V  S  G  A  A  A  Q  A  D  V  R  V  F  E   1210
             *

3841  GGAGCTCCGCGGACTCAGCGTGGACATGAGCCTGGCCGTGGAGCAGGGCGCCCCCGTGGT  3900
1211   E  L  R  G  L  S  V  D  M  S  L  A  V  E  Q  G  A  P  V  V   1230

3901  GGTCAGCGCCGCGGTGCAGACGGGCGACAACATCACGTGGACCTTCGACATGGGGGACGG  3960
1231   V  S  A  A  V  Q  T  G  D  N  I  T  W  T  F  D  M  G  D  G   1250
                             *

3961  CACCGTGCTGTCGGGCCCGGAGGCAACAGTGGAGCATGTGTACCTGCGGGCACAGAACTG  4020
1251   T  V  L  S  G  P  E  A  T  V  E  H  V  Y  L  R  A  Q  N  C   1270
                                                          *

4021  CACAGTGACCGTGGGTGCGGCCAGCCCCGCCGGCCACCTGGCCCGGAGCCTGCACGTGCT  4080
1271   T  V  T  V  G  A  A  S  P  A  G  H  L  A  R  S  L  H  V  L   1290

4081  GGTCTTCGTCCTGGAGGTGCTGCGCGTTGAACCCGCCGCCTGCATCCCCACGCAGCCTGA  4140
1291   V  F  V  L  E  V  L  R  V  E  P  A  A  C  I  P  T  Q  P  D   1310

4141  CGCGCGGCTCACGGCCTACGTCACCGGGAACCCGGCCCACTACCTCTTCGACTGGACCTT  4200
1311   A  R  L  T  A  Y  V  T  G  N  P  A  H  Y  L  F  D  W  T  F   1330

4201  CGGGGATGGCTCCTCCAACACGACCGTGCGGGGGTGCCCGACGGTGACACACAACTTCAC  4260
1331   G  D  G  S  S  N  T  T  V  R  G  C  P  T  V  T  H  N  F  T   1350
                      *                                 *

4261  GCGGAGCGGCACGTTCCCCCTGGCGCTGGTGCTGTCCAGCCGCGTGAACAGGGCGCATTA  4320
1351   R  S  G  T  F  P  L  A  L  V  L  S  S  R  V  N  R  A  H  Y   1370

4321  CTTCACCAGCATCTGCGTGGAGCCAGAGGTGGGCAACGTCACCCTGCAGCCAGAGAGGCA  4380
1371   F  T  S  I  C  V  E  P  E  V  G  N  V  T  L  Q  P  E  R  Q   1390
                                        *

4381  GTTTGTGCAGCTCGGGGACGAGGCCTGGCTGGTGGCATGTGCCTGGCCCCCGTTCCCCTA  4440
1391   F  V  Q  L  G  D  E  A  W  L  V  A  C  A  W  P  P  F  P  Y   1410

4441  CCGCTACACCTGGGACTTTGGCACCGAGGAAGCCGCCCCCACCCGTGCCAGGGGCCCTGA  4500
1411   R  Y  T  W  D  F  G  T  E  E  A  A  P  T  R  A  R  G  P  E   1430

4501  GGTGACGTTCATCTACCGAGACCCAGGCTCCTATCTTGTGACAGTCACCGCGTCCAACAA  4560
1431   V  T  F  I  Y  R  D  P  G  S  Y  L  V  T  V  T  A  S  N  N   1450
                                                             *

4561  CATCTCTGCTGCCAATGACTCAGCCCTGGTGGAGGTGCAGGAGCCCGTGCTGGTCACCAG  4620
1451   I  S  A  A  N  D  S  A  L  V  E  V  Q  E  P  V  L  V  T  S   1470
                   *

4621  CATCAAGGTCAATGGCTCCCTTGGGCTGGAGCTGCAGCAGCCGTACCTGTTCTCTGCTGT  4680
1471   I  K  V  N  G  S  L  G  L  E  L  Q  Q  P  Y  L  F  S  A  V   1490
             *

4681  GGGCCGTGGGCGCCCCGCCAGCTACCTGTGGGATCTGGGGGACGGTGGGTGGCTCGAGGG  4740
1491   G  R  G  R  P  A  S  Y  L  W  D  L  G  D  G  G  W  L  E  G   1510

4741  TCCGGAGGTCACCCACGCTTACAACAGCACAGGTGACTTCACCGTTAGGGTGGCCGGCTG  4800
1511   P  E  V  T  H  A  Y  N  S  T  G  D  F  T  V  R  V  A  G  W   1530
                            *
```

```
4801  GAATGAGGTGAGCCGCAGCGAGGCCTGGCTCAATGTGACGGTGAAGCGGCGCGTGCGGGG  4860
1531   N  E  V  S  R  S  E  A  W  L  N  V  T  V  K  R  R  V  R  G   1550
                                     *
4861  GCTCGTCGTCAATGCAAGCCGCACGGTGGTGCCCCTGAATGGGAGCGTGAGCTTCAGCAC  4920
1551   L  V  V  N  A  S  R  T  V  V  P  L  N  G  S  V  S  F  S  T   1570
          *                                 *
4921  GTCGCTGGAGGCCGGCAGTGATGTGCGCTATTCCTGGGTGCTCTGTGACCGCTGCACGCC  4980
1571   S  L  E  A  G  S  D  V  R  Y  S  W  V  L  C  D  R  C  T  P   1590

4981  CATCCCTGGGGGTCCTACCATCTCTTACACCTTCCGCTCCGTGGGCACCTTCAATATCAT  5040
1591   I  P  G  G  P  T  I  S  Y  T  F  R  S  V  G  T  F  N  I  I   1610

5041  CGTCACGGCTGAGAACGAGGTGGGCTCCGCCCAGGACAGCATCTTCGTCTATGTCCTGCA  5100
1611   V  T  A  E  N  E  V  G  S  A  Q  D  S  I  F  V  Y  V  L  Q   1630

5101  GCTCATAGAGGGGCTGCAGGTGGTGGGCGGTGGCCGCTACTTCCCCACCAACCACACGGT  5160
1631   L  I  E  G  L  Q  V  V  G  G  G  R  Y  F  P  T  N  H  T  V   1650
                                                       *
5161  ACAGCTGCAGGCCGTGGTTAGGGATGGCACCAACGTCTCCTACAGCTGGACTGCCTGGAG  5220
1651   Q  L  Q  A  V  V  R  D  G  T  N  V  S  Y  S  W  T  A  W  R   1670
                                     *
5221  GGACAGGGGCCCGGCCCTGGCCGGCAGCGGCAAAGGCTTCTCGCTCACCGTGCTCGAGGC  5280
1671   D  R  G  P  A  L  A  G  S  G  K  G  F  S  L  T  V  L  E  A   1690

5281  CGGCACCTACCATGTGCAGCTGCGGGCCACCAACATGCTGGGCAGCGCCTGGGCCGACTG  5340
1691   G  T  Y  H  V  Q  L  R  A  T  N  M  L  G  S  A  W  A  D  C   1710

5341  CACCATGGACTTCGTGGAGCCTGTGGGGTGGCTGATGGTGACCGCCTCCCCGAACCCAGC  5400
1711   T  M  D  F  V  E  P  V  G  W  L  M  V  T  A  S  P  N  P  A   1730

5401  TGCCGTCAACACAAGCGTCACCCTCAGTGCCGAGCTGGCTGGTGGCAGTGGTGTCGTATA  5460
1731   A  V  N  T  S  V  T  L  S  A  E  L  A  G  G  S  G  V  V  Y   1750
             *
5461  CACTTGGTCCTTGGAGGAGGGGCTGAGCTGGGAGACCTCCGAGCCATTTACCACCCATAG  5520
1751   T  W  S  L  E  E  G  L  S  W  E  T  S  E  P  F  T  T  H  S   1770

5521  CTTCCCCACACCCGGCCTGCACTTGGTCACCATGACGGCAGGGAACCCGCTGGGCTCAGC  5580
1771   F  P  T  P  G  L  H  L  V  T  M  T  A  G  N  P  L  G  S  A   1790

5581  CAACGCCACCGTGGAAGTGGATGTGCAGGTGCCTGTGAGTGGCCTCAGCATCAGGGCCAG  5640
1791   N  A  T  V  E  V  D  V  Q  V  P  V  S  G  L  S  I  R  A  S   1810
          *
5641  CGAGCCCGGAGGCAGCTTCGTGGCGGCCGGGTCCTCTGTGCCCTTTTGGGGGCAGCTGGC  5700
1811   E  P  G  G  S  F  V  A  A  G  S  S  V  P  F  W  G  Q  L  A   1830

5701  CACGGGCACCAATGTGAGCTGGTGCTGGGCTGTGCCCGGCGGCAGCAGCAAGCGTGGCCC  5760
1831   T  G  T  N  V  S  W  C  W  A  V  P  G  G  S  S  K  R  G  P   1850
             *
5761  TCATGTCACCATGGTCTTCCCGGATGCTGGCACCTTCTCCATCCGGCTCAATGCCTCCAA  5820
1851   H  V  T  M  V  F  P  D  A  G  T  F  S  I  R  L  N  A  S  N   1870
                                                       *
5821  CGCAGTCAGCTGGGTCTCAGCCACGTACAACCTCACGGCGGAGGAGCCCATCGTGGGCCT  5880
1871   A  V  S  W  V  S  A  T  Y  N  L  T  A  E  E  P  I  V  G  L   1890
                                     *
5881  GGTGCTGTGGGCCAGCAGCAAGGTGGTGGCGCCCGGGCAGCTGGTCCATTTTCAGATCCT  5940
1891   V  L  W  A  S  S  K  V  V  A  P  G  Q  L  V  H  F  Q  I  L   1910

5941  GCTGGCTGCCGGCTCAGCTGTCACCTTCCGCCTGCAGGTCGGCGGGGCCAACCCCGAGGT  6000
1911   L  A  A  G  S  A  V  T  F  R  L  Q  V  G  G  A  N  P  E  V   1930
```

```
6001  GCTCCCCGGGCCCCGTTTCTCCCACAGCTTCCCCCGCGTCGGAGACCACGTGGTGAGCGT  6060
1931   L  P  G  P  R  F  S  H  S  F  P  R  V  G  D  H  V  V  S  V   1950

6061  GCGGGGCAAAAACCACGTGAGCTGGGCCCAGGCGCAGGTGCGCATCGTGGTGCTGGAGGC  6120
1951   R  G  K  N  H  V  S  W  A  Q  A  Q  V  R  I  V  V  L  E  A   1970

6121  CGTGAGTGGGCTGCAGATGCCCAACTGCTGCGAGCCTGGCATCGCCACGGGCACTGAGAG  6180
1971   V  S  G  L  Q  M  P  N  C  C  E  P  G  I  A  T  G  T  E  R   1990

6181  GAACTTCACAGCCCGCGTGCAGCGCGGCTCTCGGGTCGCCTACGCCTGGTACTTCTCGCT  6240
1991   N  F  T  A  R  V  Q  R  G  S  R  V  A  Y  A  W  Y  F  S  L   2010
           *
6241  GCAGAAGGTCCAGGGCGACTCGCTGGTCATCCTGTCGGGCCGCGACGTCACCTACACGCC  6300
2011   Q  K  V  Q  G  D  S  L  V  I  L  S  G  R  D  V  T  Y  T  P   2030

6301  CGTGGCCGCGGGGCTGTTGGAGATCCAGGTGCGCGCCTTCAACGCCCTGGGCAGTGAGAA  6360
2031   V  A  A  G  L  L  E  I  Q  V  R  A  F  N  A  L  G  S  E  N   2050
                                                                 *
6361  CCGCACGCTGGTGCTGGAGGTTCAGGACGCCGTCCAGTATGTGGCCCTGCAGAGCGGCCC  6420
2051   R  T  L  V  L  E  V  Q  D  A  V  Q  Y  V  A  L  Q  S  G  P   2070

6421  CTGCTTCACCAACCGCTCGGCGCAGTTTGAGGCCGCCACCAGCCCCAGCCCCCGGCGTGT  6480
2071   C  F  T  N  R  S  A  Q  F  E  A  A  T  S  P  S  P  R  R  V   2090
              *
6481  GGCCTACCACTGGGACTTTGGGGATGGGTCGCCAGGGCAGGACACAGATGAGCCCAGGGC  6540
2091   A  Y  H  W  D  F  G  D  G  S  P  G  Q  D  T  D  E  P  R  A   2110

6541  CGAGCACTCCTACCTGAGGCCTGGGGACTACCGCGTGCAGGTGAACGCCTCCAACCTGGT  6600
2111   E  H  S  Y  L  R  P  G  D  Y  R  V  Q  V  N  A  S  N  L  V   2130
                                                 *
6601  GAGCTTCTTCGTGGCGCAGGCCACGGTGACCGTCCAGGTGCTGGCCTGCCGGGAGCCGGA  6660
2131   S  F  F  V  A  Q  A  T  V  T  V  Q  V  L  A  C  R  E  P  E   2150

6661  GGTGGACGTGGTCCTGCCCCTGCAGGTGCTGATGCGGCGATCACAGCGCAACTACTTGGA  6720
2151   V  D  V  V  L  P  L  Q  V  L  M  R  R  S  Q  R  N  Y  L  E   2170

6721  GGCCCACGTTGACCTGCGCGACTGCGTCACCTACCAGACTGAGTACCGCTGGGAGGTGTA  6780
2171   A  H  V  D  L  R  D  C  V  T  Y  Q  T  E  Y  R  W  E  V  Y   2190

6781  TCGCACCGCCAGCTGCCAGCGGCCGGGGCGCCCAGCGCGTGTGGCCCTGCCCGGCGTGGA  6840
2191   R  T  A  S  C  Q  R  P  G  R  P  A  R  V  A  L  P  G  V  D   2210

6841  CGTGAGCCGGCCTCGGCTGGTGCTGCCGCGGCTGGCGCTGCCTGTGGGGCACTACTGCTT  6900
2211   V  S  R  P  R  L  V  L  P  R  L  A  L  P  V  G  H  Y  C  F   2230

6901  TGTGTTTGTCGTGTCATTTGGGGACACGCCACTGACACAGAGCATCCAGGCCAATGTGAC  6960
2231   V  F  V  V  S  F  G  D  T  P  L  T  Q  S  I  Q  A  N  V  T   2250
                                                             *
6961  GGTGGCCCCCGAGCGCCTGGTGCCCATCATTGAGGGTGGCTCATACCGCGTGTGGTCAGA  7020
2251   V  A  P  E  R  L  V  P  I  I  E  G  G  S  Y  R  V  W  S  D   2270

7021  CACACGGGACCTGGTGCTGGATGGGAGCGAGTCCTACGACCCCAACCTGGAGGACGGCGA  7080
2271   T  R  D  L  V  L  D  G  S  E  S  Y  D  P  N  L  E  D  G  D   2290

7081  CCAGACGCCGCTCAGTTTCCACTGGGCCTGTGTGGCTTCGACACAGAGGGAGGCTGGCGG  7140
2291   Q  T  P  L  S  F  H  W  A  C  V  A  S  T  Q  R  E  A  G  G   2310

7141  GTGTGCGCTGAACTTTGGGCCCCGCGGGAGCAGCACGGTCACCATTCCACGGGAGCGGCT  7200
2311   C  A  L  N  F  G  P  R  G  S  S  T  V  T  I  P  R  E  R  L   2330
```